PRELUCRAREA DATELOR EXPERIMENTALE
Partea 1

EXPERIMENTAL DATA PROCESSING
Part 1

Wilhelm LAURENZI
Conf.dr.ing – Universitatea TRANSILVANIA din Braşov - Facultatea de Ingineria Lemnului
Adresa/Address: B-dul Eroilor nr. 29, 500036 Braşov, România
Tel:0040 268 412921/int.192. Fax: 0040 268 415315
Email: willy@unitbv.ro

Rezumat:
Orice cercetare ştiinţifică, indiferent de complexitatea ei, conţine cercetări experimentale. Aceste cercetări experimentale trebuie făcute respectând metodele statisticii matematice. Aceste metode diferă funcţie de numărul de factori de influenţă asupra procesului care se studiază. Deoarece nu toţi cercetătorii au acces la programe specializate de prelucrare a datelor experimentale, sau aceste programe nu sunt concepute pentru toate etapele de cercetare experimentală, se prezintă unele metode statistice de experimentare pentru procese monofactoriale şi un program de aplicaţie care conţine toate etapele de cercetare experimentală necesare. În această primă parte a articolului, se prezintă baza teoretică a programului de aplicaţie realizat în mediul de programare Delphi.

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

Abstract:
Every scientific research, of any complexity, is based on experimental researches. This researches must be done in accordance with the methods of mathematical statistics. These methods can differ based on the number of the influence factors on the studied process. Since not all researchers have direct access to specialized programs for processing experimental data, or these programs have not been conceived for all the stages of experimental researches, the aim of this paper is to describe a few statistical methods for experimenting with monofactorial processes, and also an application covering all the necessary stages of experimental researches. This first part of the article presents the theoretical base of a applicative program conceived in Delphi.

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

INTRODUCERE

Conform spuselor lui Anton Pann, "ştiinţa este o uşă a cărei cheie este cercetarea". Cercetarea ştiinţifică reprezintă un proces continuu de acumulare de cunoştinţe asupra unor fenomene, obiecte, procese etc. (Fig. 1). De obicei, cercetarea pleacă de la o idee, se stabilesc ipoteze preliminare, se planifică experimentele, se analizează rezultatele, se stabilesc alte ipoteze mai exacte şi se reia procesul de cercetare.

INTRODUCTION

Anton Pann said that "science is a door and research is the key". Scientific research is a continuous process of gathering knowledge about phenomena, objects, processes etc. (Fig. 1). Usually, a research starts with an idea and goes on to establish preliminary hypotheses, planning experiments, analyzing results, establish more precise hypotheses and resume the process of research.

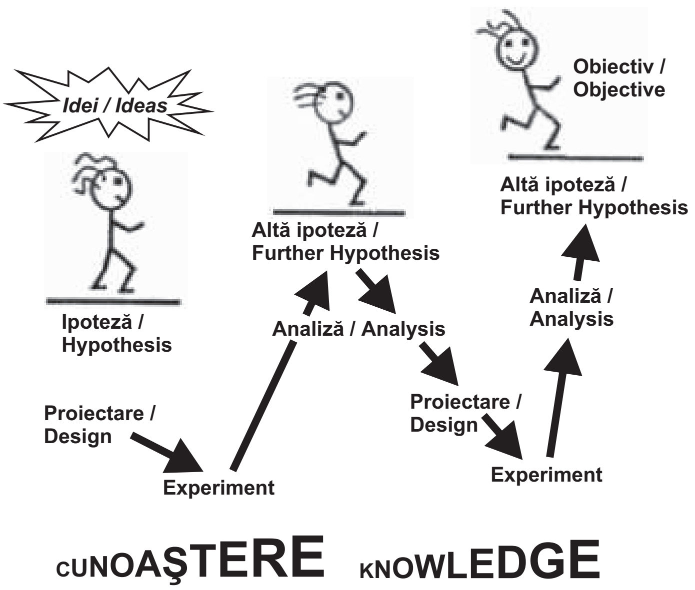

Fig. 1.
Etapele procesului de cercetare ştiinţifică /
Stages of scientific research.

Procesul de cercetare ştiinţifică se desfăşoară în general respectând următoarele etape:

•Stabilirea problemei de cercetat

•Efectuarea unui studiu bibliografic de specialitate

•Formularea unor ipoteze

•Stabilirea metodologiei de experimentare şi proiectarea experimentelor

•Efectuarea experimentelor şi achiziţia datelor

•Analiza datelor experimentale

•Interpretarea rezultatelor

•Formularea concluziilor

După cum se poate observa din cele prezentate mai sus, rezultă că experimentele au o importanţă foarte mare într-o cercetare ştiinţifică. În practică, în multe cazuri experimentele se desfăşoară după metode mai mult sau mai puţin ştiinţifice. Eliminarea arbitrariului din cercetările experimentale se poate face prin utilizarea cercetării statistice. Această cercetare cuprinde toate operaţiile de culegere, sistematizare, grupare, prelucrare, analiză şi interpretare a datelor şi informaţiilor necesare pentru cunoaşterea fenomenelor şi proceselor. Utilizarea cercetării statistice, pe lângă analiza ştiinţifică a datelor experimentale oferă posibilitatea implementării metodelor utilizate în programe de calculator.

OBIECTIVELE CERCETARII

Realizarea cercetărilor experimentale presupune cunoaşterea metodelor de cercetare statistică experimentală şi a mijloacelor (soft-uri) cu care aceste cercetări pot fi efectuate. Având în vedere că cercetările experimentale se realizează în cadrul unor procese mono- sau multifactoriale, metodele şi mijloacele de experimentare diferă funcţie de numărul de factori. Pentru a veni în întâmpinarea tinerilor cercetători, masteranzi şi doctoranzi, autorul şi-a propus realizarea unei serii de articole care să prezinte atât teoretic cât şi practic metodele statistice de prelucrare a datelor experimentale. Prezentul articol este dedicat cercetărilor experimentale ***monofactoriale*** şi cuprinde în prima parte prezentarea detaliată a metodelor statistico-matematice specifice, iar în a doua parte un program de aplicaţie care are la bază aceste metode. Metodele de cercetare statistică sunt prezentate în extenso pentru a putea fi utilizate de către studenţi sau de către alţi specialişti care doresc să-şi facă propriile aplicaţii în Excel sau în alte limbaje de programare.

METODA DE LUCRU

Ce este un experiment? O definiţie sugestivă ar fi următoarea: ”Un experiment (lat.: ex-+-periri, "despre încercare"), este un set de observaţii desfăşurate în contextul rezolvării unei anumite probleme sau chestiuni, pentru a sprijini sau falsifica o ipoteză sau cercetare privitoare la un fenomen”. (http://ro.wikipedia.org/wiki/Experiment). Deoarece fenomenele au loc în cadrul unor procese rezultă că cercetările experimentale trebuie să ţină cont de faptul că asupra proceselor acţionează unul sau mai mulţi factori de influenţă. Aceste influenţe pot fi

The process of scientific research usually goes through the following stages:

•Establishing the object of research

•Conducting bibliographic research

•Formulating hypotheses

•Establishing experimental methodology and planning experiments

•Conducting experiments and gathering data

•Analyzing experimental data

•Evaluating the results

•Formulating the consequences

As can be seen above, experiments are a major part of scientific research. In practice, experiments are often conducted through more or less scientific methods. Randomness can be excluded from experimental research through the use of statistical research. This type of research encompasses all operations necessary for gathering, systemizing, grouping, processing, analyzing and evaluating the necessary data and information for understanding phenomena and processes. Along with the scientific analysis of experimental data, the use of statistical analysis enables the implementation of the applied methods in computer programs.

OBJECTIVES

Experimental research requires prior knowledge of the methods of experimental statistical research and of the instruments (software programs) needed in the process. Given that experimental research is conducted in mono- or multi-factorial processes, the experimental methods and means can also differ based on the number of factors involved. In order to assist young researchers, Master- and PhD-students, the author proposed himself to generate a serie of articles, to present both theoretically and practically the methods for statistical processing of the experimental data. This article is dedicated to ***mono-factorial*** experimental research and it describes in detail within the first part a few specific statistical-mathematical methods, and within the second part an applicative program based on these methods. The methods of statistical research are featured extensively so as to enable students or other specialists who wish to develop their own applications in Excel or other programming languages.

METHOD

What is an experiment? A suggestive definition can be found on the internet: ”An experiment (lat.: ex-+-periri, "to try out"), is a set of observations made to solve certain problems or issues, in order to corroborate or discard a hypothesis or research matter regarding a certain phenomenon” (http://ro.wikipedia.org/wiki/Experiment). Since phenomena manifest themselves as processes, experimental research must take into account the fact that processes are influenced by one or more factors. These influences can be expressed through the following formula:

exprimate prin următoarea relație matematică:

$$y = f(x_1, x_2, x_3, ..., x_n) \quad (1)$$

unde: y – reprezintă variabila de dependență; x_1, x_2, x_3,…, x_n – variabilele independente numite şi factori de influență.

În practică există multe cazuri în care se studiază influența unui singur factor asupra procesului, cum ar fi în cazul studiului preciziei de prelucrare, a calității produselor, a rezistenței îmbinărilor etc.

Pentru realizarea experimentelor în aceste cazuri se parcurg următoarele etape:

- planificarea experimentelor;
- culegerea datelor experimentale;
- testarea datelor experimentale;
- interpretarea datelor experimentale.

PLANIFICAREA EXPERIMENTELOR

Planificarea experimentelor se face prin stabilirea numărului de experiențe (mărimea eşantionului) şi a numărului probei extrase din lotul (populația) care se studiază (Biji1979):

a) **Numărul de experiențe** (mărimea eşantionului) se poate calcula pornind de la eroarea limită maximă admisă (2):

$$n = \frac{\dfrac{t_\alpha^2 \cdot p(1-p)}{\Delta_x^2}}{1 + \dfrac{\dfrac{t^2 \cdot p \cdot (1-p)}{\Delta_x^2} - 1}{N}} \quad (2)$$

unde: n – numărul de probe din eşantion; N – numărul probelor din lot; *t* – variabila de testare Student, pentru o precizie de 5%, *t = 1.96 (pentru alte precizii se poate consulta Mihailescu 1984)*; p – proporția componentelor din eşantion care poseda caracteristica cercetată (p = 0,5); Δ_x – eroarea limită admisă (de exemplu 0,05 = 5%);

b) **Stabilirea numărului probei** care se extrage din lot se poate face prin mai multe metode printre care cea mai utilizată este metoda generării numerelor aleatoare. În acest sens se generează un număr aleator cuprins între 1 şi N, unde N reprezintă numărul de probe dintr-un lot. În urma aplicării relațiilor matematice (2-6) se generează planurile de experimentare cu o singură experiență sau mai multe experiențe (maxim 5) pentru fiecare probă. Pentru exemplificare s-a considerat, un lot de N = 30 de probe pentru care numărul de experiențe necesare pentru o precizie de 5% este de n = 28 de experiențe (Tabelul 1 sau 2).

$$y = f(x_1, x_2, x_3, ..., x_n) \quad (1)$$

where y – dependent variable; x_1, x_2, x_3,…, x_n – independent variables, also known as factors of influence.

In practice, it is common to study the influence of a single factor on the process, such as the study of processing precision, of product quality, of compound resistance etc.

In order to carry out experiments in these cases, the following steps must be followed:

- planning experiments;
- collecting experimental data;
- testing experimental data;
- interpretation the results.

PLANNING EXPERIMENTS

Planning an experiment involves establishing the number of tests (sample size) and the number of samples extracted from the analyzed population (Biji 1979):

a) **The number of tests** (sample size) can be calculated starting from the limits of permissible errors (2):

$$n = \frac{\dfrac{t_\alpha^2 \cdot p(1-p)}{\Delta_x^2}}{1 + \dfrac{\dfrac{t^2 \cdot p \cdot (1-p)}{\Delta_x^2} - 1}{N}} \quad (2)$$

where: n – number of samples extracted from the population; N – size of population; *t* – Student test variable, for a precision of 5%, *t = 1.96. For other precision values consult Table 3.10 from [Mihailescu, 1984]*; p – the proportion of components of the sample with the same characteristics (p = 0.5); Δ_x – admissible error limit (i. e. 0.05 = 5%).

b) **The number of samples** extracted from the population can be established through several methods, the most common being the method of generating random numbers. This method consists of generating a random number between 1 and N, with N being the size of the population. The application of the corresponding mathematical formula (2) leads to generating experimental plans for a single test or for more than one test (maximum 5) for each. To exemplify this, we will consider a population of N = 30, where the number of tests calculated with (2) for 5% precision being n = 28 (Table 1 or 2).

Tabelul 1 / Table 1

Planul de experimentare (un singur experiment) / Experimental plan (single test)

Nr.crt. / Ref.No.	Numărul probei extrase (valoare aleatoare) / Number of extracted sample (random value)	Valoare măsurată / Measured value
1	3	
2	12	
	…	
n	25	

Tabelul 2 / Table 2

Planul de experimentare (experiment repetat) / Experimental plan (repeated tests)

Nr.crt. / Ref.No.	Numărul probei extrase (valoare aleatoare) / Number of extracted sample (random value)	Valori măsurate (1 – 5 valori) / Measured values (1 – 5 values)				Media valorilor măsurate / Average of measured value
1	3			.....		
2	12			.....		
…	…	…	…	.....		…
n	25			…..		

După generarea planurilor de experimentare se trece la etapa următoare, respectiv culegerea datelor experimentale.

CULEGEREA DATELOR

Culegerea datelor experimentale se face conform planurilor de experimentare prezentate în Tabelul 1 sau 2. Mai întâi se extrag probele care vor fi măsurate. Măsurarea se face folosind metodele clasice sau prin înregistrare cu aparate de măsură prevăzute cu această facilitate sau pe calculator cu ajutorul plăcilor de achiziție de date. Indiferent de metoda de măsurare utilizată se completează unul din planurile de experimentare de mai jos cu valorile măsurate (Tabelul 3 sau 4).

After generating the experimental plan we can pass on to the next level, *i.e.* collecting the experimental data.

COLLECTING EXPERIMENTAL DATA

Experimental data can be collected according to the experimental plan presented in Table 1 or 2. First, we extract the sample we are about to measure. Measuring can be carried out with classical methods or with specially equipped measuring devices or on computer through data acquisition boards. Irrespective of the measuring technique you choose to employ, one of the experimental plan below must be filled out with the measured values (Table 3 or 4).

Tabelul 3 / Table 3

Planul de experimentare completat cu date experimentale (un singur experiment)/ Experimental plan, filled out with experimental data (single test)

Nr.crt. / Ref.No.	Numărul probei extrase (valoare aleatoare) / Number of extracted sample (random value)	Valoare măsurată / Measured value
1	3	19.3
2	12	18.8
…	…	…
28	25	18.9

Tabelul 4/ Table 4

Planul de experimentare completat cu date experimentale (cinci experimente) / Experimental plan, filled out with experimental data (five tests)

Nr.crt. / Ref.No.	Numărul probei extrase (valoare aleatoare) / Number of extracted sample (random value)	Valori măsurate / Measured values					Media valorilor măsurate / Average of measured values
1	3	19.3	19.1	19.4	18.7	18.9	19.08
2	12	18.8	18.7	19.3	19.0	19.1	18.98
…	…	…	…	…	…	…	...
28	25	18.9	18.8	19.4	19.3	19.4	19.16

TESTAREA DATELOR EXPERIMENTALE

După culegerea datelor experimentale acestea sunt prelucrate (http://www.labsmn.pub.ro/academic), respectiv se calculează mediile valorilor măsurate dacă este cazul (Tabelul 4), după care acestea sunt supuse unor teste statistice pentru:

- identificarea valorilor afectate de erori aberante

TESTING EXPERIMENTAL DATA

After collecting the experimental data, they must be processed. Firstly, the average measured values are calculated, if necessary (Table 4). Secondly, the data are put through a series of statistical tests in order to:

şi eliminarea acestora din eşantion;

- verificarea caracterului aleator al datelor (testul Young);
- verificarea normalității distribuției datelor experimentale.

După eliminarea valorilor aberante, verificarea caracterului aleator şi a normalității de repartiție se poate trece la modelarea datelor experimentale în vederea obținerii modelelor matematice care descriu cât mai bine procesul monofactorial y=f(x). Modul de obținere a modelelor matematice monofactoriale şi studiul influenței factorului studiat asupra procesului vor fi prezentate în partea a doua articolului.

IDENTIFICAREA DATELOR AFECTATE DE ERORI ABERANTE

Identificarea datelor afectate de erori aberante se realizează cu **testul Chauvenet** (http://www.labsmn.pub.ro/academic) sau **testul Romanovski** (Mihăilescu 1984). Datele afectate de erori aberante sunt de obicei valorile minime sau maxime ale şirului de date experimentale. O valoare este considerată aberantă dacă este îndeplinită condiția testului Chauvenet (relația 3):

$$|x_i - \bar{x}| > z \cdot \sigma \quad (3) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4)$$

unde: x_i – reprezintă valoarea afectată de erori aberante; $\bar{x}$ – media aritmetică a întregului şir de valori; z – o valoare calculată cu relația (5) şi (6); σ – abaterea medie standard a datelor experimentale (relația 7):

$$z = \frac{0.435 - 0.862 \cdot a}{1 - 3.604 \cdot a + 3.213 \cdot a^2}\ ;(5) \qquad a = \frac{2 \cdot n - 1}{4 \cdot n}\ ;(6)$$

$$\sigma = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}\ ;(7)$$

unde: n – numărul de valori ale şirului de date experimentale; x_i – şirul de date experimentale: x_1, x_2, ..., x_n; $\bar{x}$ – media aritmetică.

După testul Romanovski o valoare experimentală este considerată aberantă dacă se îndeplineşte condiția (8):

$$t > t_{1-\frac{\alpha}{2}} \quad (8) \qquad t = \frac{|x_e - \bar{x}|}{\sqrt{\frac{s^2 \cdot n}{n-1}}} \quad (9) \qquad s^2 = \frac{\sum_{i=1}^{n-1}(x_i - \bar{x})}{n-1} \quad (10)$$

unde: t – variabila de testare care se calculează cu relația (9), (10); $t_{1-\frac{\alpha}{2}}$ – variabila critică de testare Student cu n-1 grade de libertate; x_e – valoare afectată de erori aberante. După depistarea valorilor aberante acestea se elimină din şir.

VERIFICAREA CARACTERULUI ALEATOR AL DATELOR

Verificarea caracterului aleator al datelor experimentale se poate face cu ajutorul **testului Young**. Un şir de date experimentale are un caracter aleator dacă este îndeplinită următoare condiție:

- identify values affected by aberrant errors and eliminate them from the sample. Systematical errors can only be eliminated by replacing the measuring system and re-conducting the experiments;
- verify aleatory of data (Young test);
- verify the normality of the distribution of experimental data.

After eliminating aberrant values, verifying aleatory of data and normality of the distribution, we can model the experimental data in order to obtain those mathematical models which best describe the monofactorial process y=f(x). Obtaining monofactorial mathematical models and studying the influence of the researched factor on the process will be presented in the up-following part of the article.

IDENTIFYING DATA AFFECTED BY ABERRANT ERRORS

Identifying data affected by aberrant errors can be accomplished by applying the **Chauvenet test** (http://www.labsmn.pub.ro/academic) or the **Romanovski test** (Mihailescu 1984). Affected data are usually the minimal or maximal values in an array of experimental data. A value is considered to be affected if the Chauvenet condition (3) is true:

$$|x_i - \bar{x}| > z \cdot \sigma \quad (3) \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4)$$

where: x_i – value affected by aberrant errors; $\bar{x}$ – arithmetic average of the entire array of values; z – value calculated with the formula (5) and (6); σ – mean standard deviation of experimental data (7):

$$z = \frac{0.435 - 0.862 \cdot a}{1 - 3.604 \cdot a + 3.213 \cdot a^2}\ ;(5 \qquad a = \frac{2 \cdot n - 1}{4 \cdot n}\ ;(6)$$

$$\sigma = \sqrt{\frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}\ ;(7)$$

where: n – number of values in the array of experimental data; x_i – array of experimental data: x_1, x_2, ..., x_n; $\bar{x}$ – arithmetic average.

After Romanovski test, an experimental data is considered affected by aberrant errors when the condition (8) is fulfilled:

$$t > t_{1-\frac{\alpha}{2}} \quad (8) \qquad t = \frac{|x_e - \bar{x}|}{\sqrt{\frac{s^2 \cdot n}{n-1}}} \quad (9) \qquad s^2 = \frac{\sum_{i=1}^{n-1}(x_i - \bar{x})}{n-1} \quad (10)$$

where: t – the test variable is calculated with the formula (9) and (10); $t_{1-\frac{\alpha}{2}}$ – Student critical test variable with n-1 degree of freedom; x_e – value affected by aberrant errors. After detecting aberrant values, they are rejected from the experimental data.

VERIFYING THE ALEATORY CHARACTER OF DATA

The aleatory character of experimental data can be verified with the **Young test**. An string of experimental data is aleatory if the following condition is true:

$$vci < m < vcs \quad (11)$$

unde: vci – valoare critică inferioară (12); m – coeficient testare (14); vcs - valoare critică superioară care se calculează cu relația (13)

$$vci = \begin{cases} 0.491 + 0.081 \cdot n - 0.003 \cdot n^2, \; pentru - \alpha = 0.95 \\ \dfrac{192.883 + 1.269 \cdot n^{2.336}}{411.427 + n^{2.336}}, \; pentru - \alpha = 0.99 \end{cases} \quad (12)$$

$$vcs = \begin{cases} 3.317 - 1.057 \cdot e^{-8.919 \cdot n^{-0.941}}, \; pentru - \alpha = 0.95 \\ 3.484 - 0.882 \cdot e^{-33.574 \cdot n^{-1.399}}, \; pentru - \alpha = 0.99 \end{cases} \quad (13)$$

$$m = \frac{\delta^2}{\sigma^2} \quad (14) \qquad \delta^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \quad (15)$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (16)$$

unde: n – numărul de valori ale şirului de date experimentale; x_i – şirul de date experimentale: x_1, x_2, ..., x_n; σ – dispersia datelor experimentale; $\bar{x}$ – media aritmetică; α – nivelul de încredere.

VERIFICAREA NORMALITĂȚII DISTRIBUȚIEI DATELOR EXPERIMENTALE

Verificarea normalității repartizării datelor experimentale se poate face cu testele Kolmogorov-Smirno, Shapiro-Wilk şi Lillefors (Statistica 2009). După Biji (1979) şi Mihailescu (1984) verificarea se face prin:

a) reprezentarea datelor experimentale sub formă de histogramă;

b) verificarea diferenței dintre media eşantionului şi valoarea mediană;

c) verificarea mărimii coeficientului de boltire şi a excesului;

d) verificarea mărimii coeficientului de asimetrie.

a) Reprezentarea datelor experimentale sub formă de histogramă. O repartiție este considerată normală dacă histograma valorilor experimentale are un singur vârf (punct de maxim).

Reprezentarea datelor experimentale sub formă de histogramă presupune:

- împărțirea datelor experimentale în clase. Numărul de clase k se calculează cu:

$$k = 1 + 3.222 \cdot \lg n \quad (17)$$

unde: n – numărul de date experimentale. Se recomandă ca numărul de clase să îndeplinească condiția: $10 \le k \le 25$ (18)

- stabilirea amplitudinii unei clase cu relația:

$$\delta = \frac{x_{max} - x_{min}}{k} \quad (19)$$

unde: x_{min} – valoarea minimă a şirului de valori experimentale; x_{max} – valoarea maximă a şirului de valori experimentale; k – numărul de clase

- stabilirea valorilor inferioare $x_{i\,inf}$ şi superioare $x_{i\,sup}$ ale claselor: $x_{i\inf} = x_i$, $x_{i\sup} = x_i + \delta$, i=1,n-1(20)

- determinarea valorilor centrale ale claselor

$$xc_i = \frac{x_{i\sup} - x_{i\inf}}{2} \quad (21)$$

$$vci < m < vcs \quad (11)$$

where: vci – critical inferior value (12); m – testing coefficient (14); vcs – critical superior value, calculated with the formula (13)

$$vci = \begin{cases} 0.491 + 0.081 \cdot n - 0.003 \cdot n^2, \; pentru - \alpha = 0.95 \\ \dfrac{192.883 + 1.269 \cdot n^{2.336}}{411.427 + n^{2.336}}, \; pentru - \alpha = 0.99 \end{cases} \quad (12)$$

$$vcs = \begin{cases} 3.317 - 1.057 \cdot e^{-8.919 \cdot n^{-0.941}}, \; pentru - \alpha = 0.95 \\ 3.484 - 0.882 \cdot e^{-33.574 \cdot n^{-1.399}}, \; pentru - \alpha = 0.99 \end{cases} \quad (13)$$

$$m = \frac{\delta^2}{\sigma^2} \quad (14) \qquad \delta^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2 \quad (15)$$

$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 \quad (16)$$

where: n – number of values in the array of experimental data; x_i – array of experimental data: x_1, x_2, ..., x_n; σ – dispersion of experimental data; $\bar{x}$ – arithmetic average; α – trust level.

VERIFYING THE NORMALITY OF THE EXPERIMENTAL DATA DISTRIBUTION

The normality of the experimental data distribution can be verified with the Kolmogorov-Smirnov test, the Shapiro-Wilk test and the Lillefors test (Statistica 2009). After Biji (1979) and Mihăilescu (1984) the normality is verified through:

a) representing experimental data as a histogram;

b) verifying the difference between the sample average and the mean value;

c) verifying the value of the kurtosis coefficient;

d) verifying the value of the skewness coefficient.

a) Representing experimental data as a histogram. A distribution is considered normal if the histogram of experimental values shows a single peak (maximum point). Representing experimental data as a histogram requires:

- dividing the experimental data into several classes. The number of classes can be calculated with the formula:

$$k = 1 + 3.222 \cdot \lg n \quad (17)$$

where: n – number of experimental data. It is recommendable that the number of classes fulfils the following condition: $10 \le k \le 25$ (18)

- establishing the amplitude of a class with the formula:

$$\delta = \frac{x_{max} - x_{min}}{k} \quad (19)$$

where: x_{min} – lowest value of the experimental data; x_{max} – highest value of the experimental data; k – number of classes.

- establishing the inferior $x_{i\,inf}$ and superior values $x_{i\,sup}$ of the classes: $x_{i\inf} = x_i$,$x_{i\sup} = x_i + \delta$, i=1,n-1(20)

- defining the central values of the classes

$$xc_i = \frac{x_{i\sup} - x_{i\inf}}{2} \quad (21)$$

b) Verificarea diferenței dintre media eşantionului şi valoarea mediana. O repartiție este considerată normală dacă diferența dintre media eşantionului $\bar{x}$ şi valoarea mediană Me a acestuia este nulă: $\bar{x} - Me = 0$. Valoarea medianei şi a mediei se determinată cu relațiile:

$$Me = \begin{cases} \dfrac{x^{\frac{n}{2}} + x^{\frac{n}{2}+1}}{2}, pentru\_n\_par \\ x^{\frac{n+1}{2}}, pentru\_n\_impar \end{cases} \quad (22)$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (23)$$

unde: n – numărul de valori a şirului de date experimentale.

c) Verificarea mărimii coeficientului de boltire. O repartiție este considerată normală dacă coeficientul de boltire β (24) are valoarea 3 şi excesul E (27) are valoarea 0:

$$\beta = \frac{\bar{\mu}_4}{\sigma^4} = 3 \quad (24)$$

unde: $\bar{\mu}_4$ reprezintă momentul centrat de ordinul 4, determinat cu relația:

$$\bar{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 \quad (25)$$

σ - abaterea standard care este determinată de această dată cu relația:

$$\sigma = \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (26) \qquad E = \beta - 3 = 0 \quad (27)$$

d) Verificarea mărimii coeficientului de asimetrie. O repartiție este simetrică dacă coeficientul de asimetrie As = 0. Coeficientul de asimetrie se calculează cu:

$$As = \frac{\bar{\mu}_3}{\sigma^3} = 0 \quad (28)$$

unde: $\bar{\mu}_3$ reprezintă momentul centrat de ordinul 3, determinat cu relația:

$$\bar{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \quad (29)$$

Dacă cu testele prezentate mai sus nu se verifică normalitatea repartiției datelor experimentale se apelează la **testul Massey** (a) sau χ^2 (b). Alegerea tipului de test se face funcție de volumul eşantionului de date experimentale. Astfel pentru eşantioane cu 8 – 50 date experimentale se alege testul Massey şi pentru eşantioane mai mari de 50 date experimentale se alege testul χ^2.

a) Testul Massey. Conform testului Massay o repartiție este normală dacă se respectă condiția:

$$d_{max} < d_{critic} \quad (30)$$

unde d_{max} – reprezintă valoarea maximă a şirului de valori d_i (31); d_{critic} – se calculează cu relațiile (36);

$$d_i = |F_i - \varphi_i - 0.5|; i = 1, n \quad (31)$$

unde: F_i - frecvențele relative cumulate:

$F_i = \dfrac{n_i}{n}; i = 1, n$ (32); n_i – numărul de valori

b) Verifying the difference between the sample average and the mean value. A distribution is considered normal if the difference between the sample average $\bar{x}$ and the mean value Me is zero: $\bar{x} - Me = 0$. The mean value Me and the average value $\bar{x}$ are determined with the formula:

$$Me = \begin{cases} \dfrac{x^{\frac{n}{2}} + x^{\frac{n}{2}+1}}{2}, pentru\_n\_par \\ x^{\frac{n+1}{2}}, pentru\_n\_impar \end{cases} \quad (22)$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (23)$$

where: n – number of experimental data.

c) Verifying the value of the kurtosis coefficient. A distribution is considered normal if the kurtosis coefficient β (24) is 3 and the excess E (27) is 0:

$$\beta = \frac{\bar{\mu}_4}{\sigma^4} = 3 \quad (24)$$

where: $\bar{\mu}_4$ represents the centered moment of fourth order, determined with the formula:

$$\bar{\mu}_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4 \quad (25)$$

σ - standard deviation, determined with the formula:

$$\sigma = \sqrt{\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (26) \qquad E = \beta - 3 = 0 \quad (27)$$

d) Verifying the value of the skewness coefficient. A distribution is considered symmetrically if skewness coefficient As (28) is 0. The skewness coefficient is calculated with:

$$As = \frac{\bar{\mu}_3}{\sigma^3} = 0 \quad (28)$$

where: $\bar{\mu}_3$ represents the centered moment of third order, determined with the formula:

$$\bar{\mu}_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3 \quad (29)$$

If the tests presented above do not verify the normal distribution of experimental data, then either the **Massey Test** (a) or the χ^2 Test (b) must be taken. Choosing which test is used depending on the size of the samples of experimental data. The Massey test is designed for samples of 8 – 50 experimental data, while the χ^2 test should be used for samples of more than 50 experimental data.

a) The Massey Test. According to the Massey test, a distribution is normal if the condition (28) is true:

$$d_{max} < d_{critic} \quad (30)$$

where d_{max} – maximum value of array d_i (29); d_{critic} – is calculated with (36);

$$d_i = |F_i - \varphi_i - 0.5|; i = 1, n \quad (31)$$

where: F_i – cumulated relative frequencies:

$F_i = \dfrac{n_i}{n}; i = 1, n$ (32); n_i – number of experimental experimentale din fiecare clasă; φ_i – valorile densității de repartiție:

$$\varphi_i = \frac{1}{2} - \frac{e^{-\frac{y_i^2}{2}}}{\sqrt{2\cdot\pi}}\left(0.4361\cdot t_i - 0.1202\cdot t_i^2 + 0.9373\cdot t_i^3\right); i = \overline{1,n} \quad (33)$$

$$y_i = \frac{x_i - \bar{x}}{\sigma}; i = 1,...,n \quad (34)$$

$$t_i = \frac{1}{1 + 0.3326\cdot y_i}; i = \overline{1,n} \quad (35)$$

$$d_{critic} = \begin{cases} 0.1851 - 0.01064\cdot n + 0.000785\cdot n^2; pentru - \alpha = 0.95 \\ 0.1408 + 0.00714\cdot n - 0.000769\cdot n^2; pentru - \alpha = 0.90 \end{cases} \quad (36)$$

unde: α - nivel de încredere.

b) Testul χ^2. Conform testului χ^2 o repartiție este normală dacă se respectă testul:

$\chi^2 > \chi^2_{critic}$ (37)

unde χ^2 – variabila test care se calculează cu:

$$\chi^2 = \sum_{i=1}^{\nu+1} \frac{n_i - n\cdot p_i}{n\cdot p_i} \quad (37)$$

$$p_i = \Phi(t_i) - \Phi(t_{i-1}), i = \overline{1,(\nu+1)} \quad (38)$$

$$\Phi(t) = \frac{1}{2} - \frac{e^{\frac{t^2}{2}}}{\sqrt{2\cdot\pi}}\left(0.4362\cdot a - 0.1202\cdot a^2 + 0.9373\cdot a^3\right) \quad (39)$$

$$a = \frac{1}{1 + 0.3326\cdot t} \quad (40) \qquad \Phi(+\infty) = 0.5 \quad (41)$$

$$\Phi(-t) = -\Phi(t) \quad (42) \qquad t_i = \frac{x_i - \bar{x}}{\sigma}; i = \overline{1,(\nu+1)} \quad (43)$$

unde: t_i – valoare calculată pentru fiecare clasă; ν - numărul gradelor de libertate = k – 1; k – numărul de clase, $k = 1 + 3.222\cdot \lg n$; x_i – valoarea superioară a fiecărei clase; χ^2_{critic} - variabila critică care se calculează cu:

$$\chi^2_{critic} = a + b\cdot\upsilon + c\cdot\nu^2 + d\cdot\nu^3 \quad (44)$$

$$a = \frac{0.2046 - 0.2032\cdot\alpha}{1 - 1.9562\cdot\alpha + 0.9563\cdot\alpha^2} \quad (45)$$

$$b = \frac{0.685 - 0.6819\cdot\alpha}{1 - 1.6662\cdot\alpha + 0.6673\cdot\alpha^2} \quad (46)$$

$$c = \frac{-0.1507 + 0.1508\cdot\alpha}{10 - 17.1239\cdot\alpha + 7.122\cdot\alpha^2} \quad (47)$$

$$d = \frac{0.07773 - 0.07928\cdot\alpha}{100 - 165.0049\cdot\alpha + 64.2821\cdot\alpha^2} \quad (48)$$

INTERPRETAREA REZULTATELOR PRELUCRĂRILOR STATISTICE

În urma prelucrării datelor experimentale monofactoriale cu metodele statistice prezentate mai sus se interpretează rezultatele astfel:

- datele experimentale aberante influențează negativ rezultatele experimentelor şi se elimină din şirul de date. Numărul acestora indică existența unor factori perturbatori în procesul de măsurare. În cazul în care se observă erori sistematice, acestea nu pot fi eliminate decât prin înlocuirea sistemului de măsurare şi refacerea experimentelor.
- datele experimentale trebuie să aibă un

values in each class; φ_i – values of distribution density:

$$\varphi_i = \frac{1}{2} - \frac{e^{-\frac{y_i^2}{2}}}{\sqrt{2\cdot\pi}}\left(0.4361\cdot t_i - 0.1202\cdot t_i^2 + 0.9373\cdot t_i^3\right); i = \overline{1,n} \quad (33)$$

$$y_i = \frac{x_i - \bar{x}}{\sigma}; i = 1,...,n \quad (34)$$

$$t_i = \frac{1}{1 + 0.3326\cdot y_i}; i = \overline{1,n} \quad (35)$$

$$d_{critic} = \begin{cases} 0.1851 - 0.01064\cdot n + 0.000785\cdot n^2; pentru - \alpha = 0.95 \\ 0.1408 + 0.00714\cdot n - 0.000769\cdot n^2; pentru - \alpha = 0.90 \end{cases} \quad (36)$$

where: α - trust level

b) The χ^2 Test. According to the χ^2 test, a distribution is normal if the condition (35) is true:

$\chi^2 > \chi^2_{critic}$ (37)

where χ^2 – test variable, calculated with:

$$\chi^2 = \sum_{i=1}^{\nu+1} \frac{n_i - n\cdot p_i}{n\cdot p_i} \quad (37)$$

$$p_i = \Phi(t_i) - \Phi(t_{i-1}), i = \overline{1,(\nu+1)} \quad (38)$$

$$\Phi(t) = \frac{1}{2} - \frac{e^{\frac{t^2}{2}}}{\sqrt{2\cdot\pi}}\left(0.4362\cdot a - 0.1202\cdot a^2 + 0.9373\cdot a^3\right) \quad (39)$$

$$a = \frac{1}{1 + 0.3326\cdot t} \quad (40) \qquad \Phi(+\infty) = 0.5 \quad (41)$$

$$\Phi(-t) = -\Phi(t) \quad (42) \qquad t_i = \frac{x_i - \bar{x}}{\sigma}; i = \overline{1,(\nu+1)} \quad (43)$$

where: t_i – value calculated for every class; ν - number of degrees of freedom = k – 1; k – number of classes, $k = 1 + 3.222\cdot \lg n$; x_i – superior value of every class; χ^2_{critic} - critical variable, calculated with:

$$\chi^2_{critic} = a + b\cdot\upsilon + c\cdot\nu^2 + d\cdot\nu^3 \quad (44)$$

$$a = \frac{0.2046 - 0.2032\cdot\alpha}{1 - 1.9562\cdot\alpha + 0.9563\cdot\alpha^2} \quad (45)$$

$$b = \frac{0.685 - 0.6819\cdot\alpha}{1 - 1.6662\cdot\alpha + 0.6673\cdot\alpha^2} \quad (46)$$

$$c = \frac{-0.1507 + 0.1508\cdot\alpha}{10 - 17.1239\cdot\alpha + 7.122\cdot\alpha^2} \quad (47)$$

$$d = \frac{0.07773 - 0.07928\cdot\alpha}{100 - 165.0049\cdot\alpha + 64.2821\cdot\alpha^2} \quad (48)$$

INTERPRETING THE RESULTS OF STATISTICAL PROCESSING

After processing the monofactorial experimental data with the statistical methods presents above the results are interpreted so:

- the aberrant experimental data influence negative the results of experiments and they are rejected from the data string. The number of these indicate the existence of perturbation factors in the measuring process. Systematical errors can only be eliminated if the measuring system is replaced and the experiments are repeated.
- the experimental data must have an aleatory character. If the data are not aleatory then the

caracter aleator. În cazul în care datele nu sunt aleatoare se reface experimentul.

- repartiția datelor experimentale trebuie să fie cât mai apropiată de repartiția normală. Aceasta se poate observa prin:

a) reprezentarea histogramei care trebuie să aibă un singur vârf;

b) prin compararea graficelor celor două funcții de repartiție. Funcția de repartiție a datelor reale trebuie să fie cât mai apropiată de funcția de repartiție normală.

c) prin compararea abaterii medii pătratice a datelor reale cu abaterea medie pătratică a repartiției normale care are valoarea $\sigma = 1$.

d) prin analiza mărimii intervalului de împrăştiere a datelor în jurul mediei $\bar{x} \pm 3\sigma$. Cu cât acest interval este mai mic cu atât datele sunt mai grupate în jurul mediei.

e) prin analiza coeficientul de boltire β care exprimă ascuțimea curbei de repartiție şi care trebuie să fie în jurul valorii de 3. Dacă $\beta > 3$ atunci curba este mai ascuțită şi datele sunt concentrate în jurul valorii medii.

f) prin analiza coeficientul de asimetrie As care exprimă simetria curbei de repartiție şi care trebuie să fie în jurul valorii de 0. Dacă As<>0 atunci curba este asimetrică, deplasată spre stânga sau spre dreapta față de curba de repartiție normală. În acest caz se verifică cauzele care au determinat asimetria curbei de repartiție.

CONCLUZII

Cercetările experimentale reprezintă o parte importantă în orice cercetare ştiințifică. Prin aceste cercetări se pot studia influențele unuia sau mai multor factori asupra unui proces. În funcție de numărul de factori de influență se utilizează anumite metode matematico-statistice pentru prelucrarea datelor experimentale. Aceste metode permit realizarea unor analize şi interpretări statistice exacte şi tragerea unor concluzii corecte. Având în vedere importanța cercetărilor experimentale se vor publica în serial metodele statistice de prelucrarea datelor experimentale mono- şi multifactoriale. În această lucrare sunt prezentate doar metodele pentru prelucrarea statistică a datelor experimentale monofactoriale. Acestea constituie baza teoretică şi algoritmul pentru un program de prelucrare statistică a datelor experimentale în Delphi care va fi prezentat în numărul următor al revistei PRO LIGNO. De asemenea, se vor prezenta pe baza unor exemple practice, etapele de planificare, culegere şi testare a datelor experimentale şi interpretarea rezultatelor obținute în urma prelucrărilor statistice.

experiment must be repeated.

- the distribution of the experimental data must be closely with the normal distribution. This can be observed by:

a) representing the histogram which must have a single peak;

b) comparing the graphics of both distribution functions. The distribution function of the real data be closely with the normal distribution.

c) comparing the standard deviation of the real data with the standard deviation of the normal distribution where $\sigma = 1$.

d) analyzing the spreading interval of the data around the mean value $\bar{x} \pm 3\sigma$. A closer interval means that the data are grouped around the mean.

e) analyzing the kurtosis coefficient β which express the sharpness of the distribution curve and whose value must be around 3. If $\beta > 3$ then the curve is more sharpen and the data are concentrate around the mean value.

f) analyzing the skewness coefficient As which express the symmetry of the distribution curve and whose value must be around 0. If As <> 0 then the curve is asymmetrical, placed left or right against the normal distribution. In this case the causes which determine the asymmetry of the distribution curve.

CONCLUSIONS

Experimental researches are an important part of any scientific research. Among other things, this type of research studies the influence of various factors on a process. Based on the number of these factors, certain mathematical-statistical methods are employed in order to process experimental data, thus enabling us to make exact analyses and interpretations and obtain right conclusions. Considering the importance of experimental researches several papers will be published in this journal where theoretical and practical statistic methods for mono- and multifactorial data processing will be presented. In this order, in this paper are presented only the methods for statistical processing of mono factorial experimental data. This method constitutes the theoretical base and the algorithm of a program in Delphi for statistical processing of experimental data which will be published in the up-following number of PRO LIGNO journal. Also there will be presented, on base of a practical example, the stages of planning, collecting and testing of experimental data and the interpretation of the results obtained by statistical processing.

BIBLIOGRAFIE / REFERENCES

BIJI, M. (1979). Statistică teoretică (Theoretical Statistics). Editura Didactică şi Pedagogică, Bucureşti.

MIHĂILESCU, T. (1984). Metode de cercetare operațională şi programarea calculatoarelor în Industria Lemnului (Operational Reserch Methods and Computer Programming in Wood Industry). Universitatea TRANSILVANIA din Braşov.

***http://ro.wikipedia.org/wiki/Experiment

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

***Statistica - http://www.statsoft.com/