

# MATHCAD PROGRAM FOR CHECKING RANDOMNESS OF SAMPLES, ELIMINATING OUTLIER DATA AND TESTING THE CONCORDANCE WITH THE WEIBULL DISTRIBUTION

Victoria TOMA<sup>1</sup>, Cristin-Olimpiu MORARIU<sup>2</sup>, Vladimir MĂRĂSCU-KLEIN<sup>1</sup>

<sup>1</sup>University Transilvania Braşov, Department of Economic Engineering and Production Systems, <sup>2</sup>Department of Manufacturing Engineering,  
e-mail: [vtoma@unitbv.ro](mailto:vtoma@unitbv.ro), [c.morariu@unitbv.ro](mailto:c.morariu@unitbv.ro), [klein@unitbv.ro](mailto:klein@unitbv.ro)

**Keywords:** reliability, checking randomness, eliminating outlier data, testing conformity

**Abstract:** Within this paper the MathCAD program is presented and developed for checking the times of good working of complex production equipment, with regard to the statistical modeling of the working time without failures.

The program allows:

- to check randomness of the experimental data by the number of iterations, the upper value criterion of iterations lengths;
  - to test and to eliminate outlier data by using the rank limits method;
  - to test the conformity of the sampling values with the three-parametric Weibull's distribution model.
- Finally, the paper presents a case study.

## 1. INTRODUCTION

For the main sub-assemblies of the production equipment and for their wear items [4], to determine the time evolution of the failing rate is indicated, for conceiving a correct maintenance policy.

Statistic modeling of the working time without failure is achieved by using a program, written in MathCAD, which shall provide:

- to check randomness of the experimental data by the number of iterations, the upper value criterion of iterations lengths;
- to test and to eliminate outlier data by using the rank limits method;
- to test the conformity of the sampling values with the three-parametric Weibull's distribution model.

## 2. NOTATIONS AND TERMINOLOGY

Within this article, the following notations and terminology has been used:

$F(t)$ -Weibull probability distribution;	$\alpha$ - significance level;
$B$ - shape parameter;	$1-\alpha$ - confidence level;
$\gamma$ - location parameter;	$C_{cor}$ - correlation coefficient;
$\eta$ - scale parameter;	$LR_j$ - lower rank limit of failure according to statistics of order $j$ ;
$N$ - sample volume;	$UR_j$ - upper rank limit of failure according to statistics of order $j$ ;
$Me$ - sample median;	

## 3. PROGRAM DESCRIPTION

The MathCAD program destined to check experimental data contains four parts:

1. checking randomness of the experimental data by the number of iterations, the maximal value criterion of iterations lengths;
2. testing and eliminating outlier data by using the rank limits method;

3. estimating the parameters of the Weibull distribution;
4. testing the conformity of the sampling values with the three-parametric Weibull's distribution model.

### 3.1. Checking randomness of the experimental data

Checking randomness of the experimental data is achieved by using the iterations method. The iteration represents serial numeric values, of the time of good working  $T_i$ ,  $i \in \overline{1, n}$ , which have the same feature, being lesser or higher compared to the grouping center, featured by the sample's median. The number of values forming an iteration is called as iteration length.

The value of iterations length  $K_{N,\alpha}$  is determined by relation [2]:

$$K_{N,\alpha} = \frac{\log \frac{-0.43429 \cdot N}{\log(1-\alpha)}}{\log 2} - 1 \quad (1)$$

The statistic decision concerning randomness is taken on grounds of the maximal length of the existing iteration ( $K_m$ ) such as:

- if  $K_m \leq K_{N,\alpha}$ , the hypothesis of randomness of the experimental data is accepted;
- if  $K_m \geq K_{N,\alpha}$ , the hypothesis of randomness of the experimental data is rejected.

### 3.2. Estimating the parameters of the Weibull's distribution

Statistical modeling of the working time between failures imposes to identify and to remove data outliers. Eliminating data outliers is achieved by applying specific tests for the distribution type, needing previously to adopt a theoretic distribution and to estimate specific parameters.

Weibull's distribution provides a good concordance with the experimental results, being a function of three parameters,  $\alpha$ ,  $\beta$ ,  $\gamma$ . Out of these considerations, its application field is a very large one. Random processes, ageing and wearing are covered. Consequently, for the statistical modeling of the times of working between accidental failures of the production equipment [3] the theoretic Weibull's model is adopted. Estimation of the location parameter  $\gamma$  is achieved by the method of the correlation coefficient and than, by applying the method of least squares to the resulted two-parametered Weibull's sample, the values of parameters  $\beta$ , respectively  $\eta$  are estimated.

### 3.3. Eliminating outlier data by the rank limits method

The Mathcad method identifies outliers by using the rank limits method [1]. The major advantage of this method consists in testing all values belonging to the observation sequence, not only of the extreme values.

For applying this method it is necessary:

- to determine the number of data,  $N$ , located within the observation rank;
- to adopt theoretic distribution and to estimate specific parameters;
- to fix the degree of significance  $\alpha$ .

Under these terms, for each observed value of the good working time, it shall be determined: the probability of failure  $F(T_j)$ ,  $j=1,2,...,N$  and the values of rank limits: lower rank limit  $LR_{j;N,\alpha/2}$  and upper rank limit  $UL_{j;N,1-\alpha/2}$ .

Testing the values pertaining to the observation rank is achieved in a differentiate manner. First of all, extreme values  $T_1$  and  $T_N$  of the observation rank are checked. If these don't represent outliers, it follows the testing of each  $T_j$  value, comprised between the values  $T_1$  and  $T_N$ .

If one or both  $T_1$ ,  $T_N$  values prove to be outliers, these are eliminated and the application of the method is resumed for the new formed rank, having  $N-1$ , respectively  $N-2$  observations.

The statistical decision is taken as follows:

- ☑ probability values according to time  $T_1$  and those according to time  $T_N$  are compared with the lower probability of failure, according to the first order statistics, respectively with the upper one, according to the  $N$  order statistics:

- $T_1$  is an outlier if  $F(T_1) \leq LR_{1;N;\alpha}$ ;
- $T_N$  is an outlier if  $F(T_N) \geq UR_{N;N;1-\alpha}$ ;

- ☑ probability values according to times  $T_j$  are checked if they are located within the range formed by the two failure probability limits:

- $T_j$  is an outlier if:  $F(T_j) \notin [LR_{j;N;\alpha/2}; UL_{j;N;1-\alpha/2}]$ .

At the end of the third part, the program displays the statistical decision concerning the observation values. If the decision is: "*Sampling values contain outsiders*" the position indexes of the respective data are displayed, with regard to their identification and elimination.

### 3.4. Testing the concordance between the sampling values and the Weibull distribution

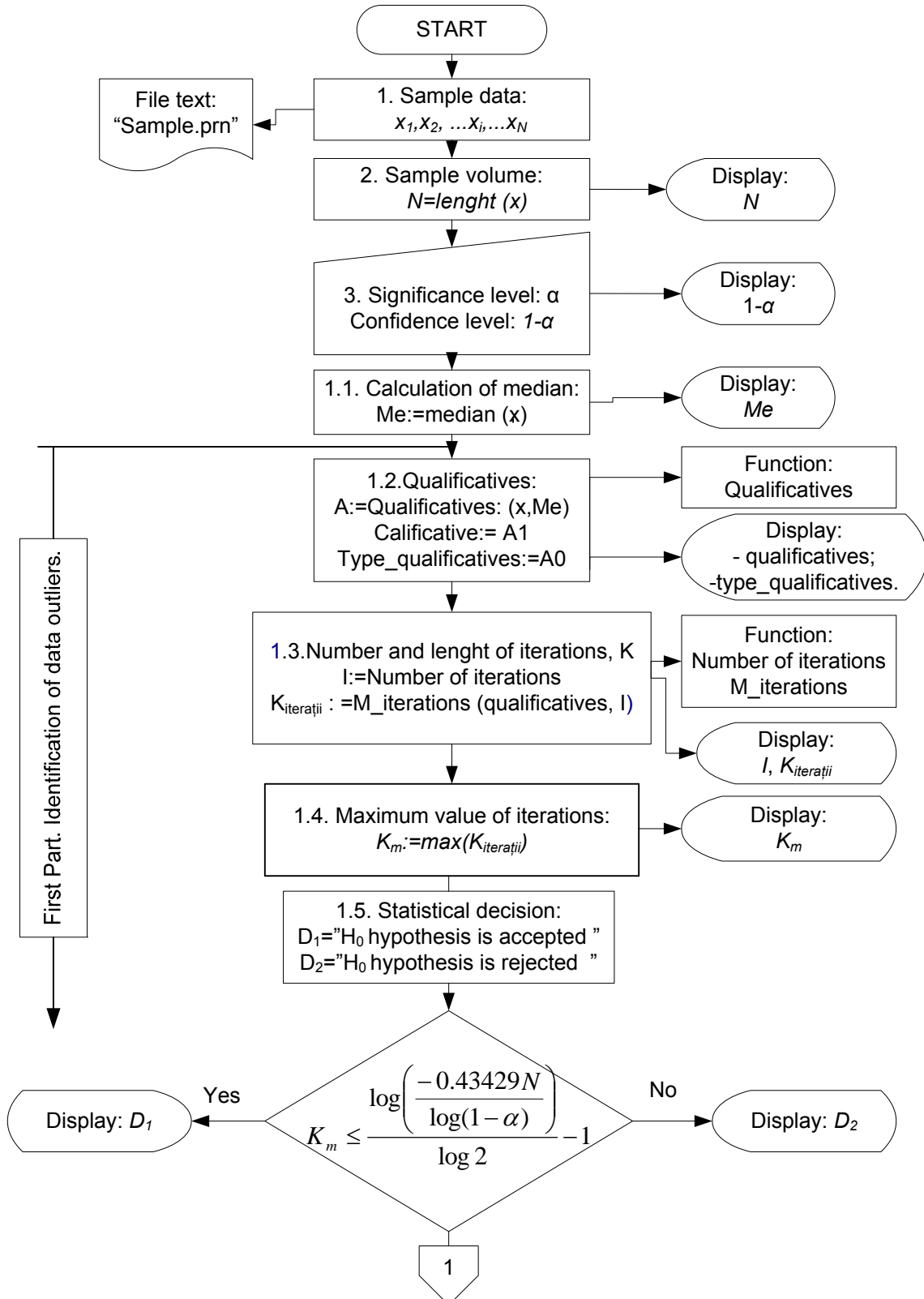
Within the experimental data processing resulting from the survey of the equipment's working, determination of the adequate statistic model is grounded because all maintenance forecasts are realized on this ground.

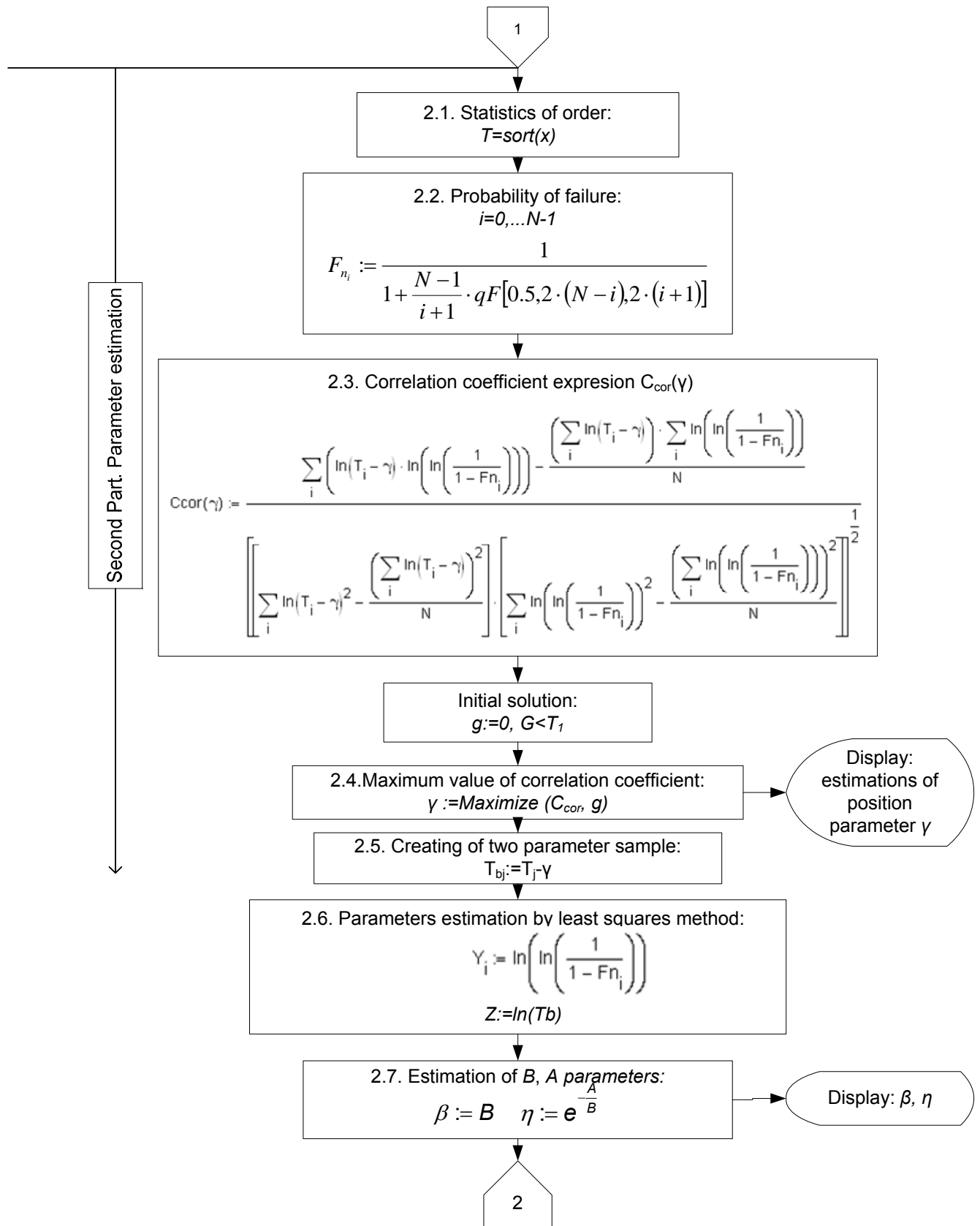
In its fourth part, the program checks if experimental data respects the hypothesis on the adopted theoretic model, by using the correlation coefficient as indicator of the concordance between experimental values  $T_1, T_2, \dots, T_N$  and the theoretic Weibull model.

The values of the correlation coefficient  $Ccor(y) \geq 0,90$  [1],[5] are showing a good correlation between the theoretic model and the observed experimental values.

In the presented Mathcad program, the statistical decision of concordance between experimental values and the theoretic ones, is taken under the terms, that:  $Ccor(y) > 0,90$ .

The logic scheme of the MathCAD program for checking the randomness of samples, eliminating outliers and testing the concordance with the Weibull model is presented in fig. 1.





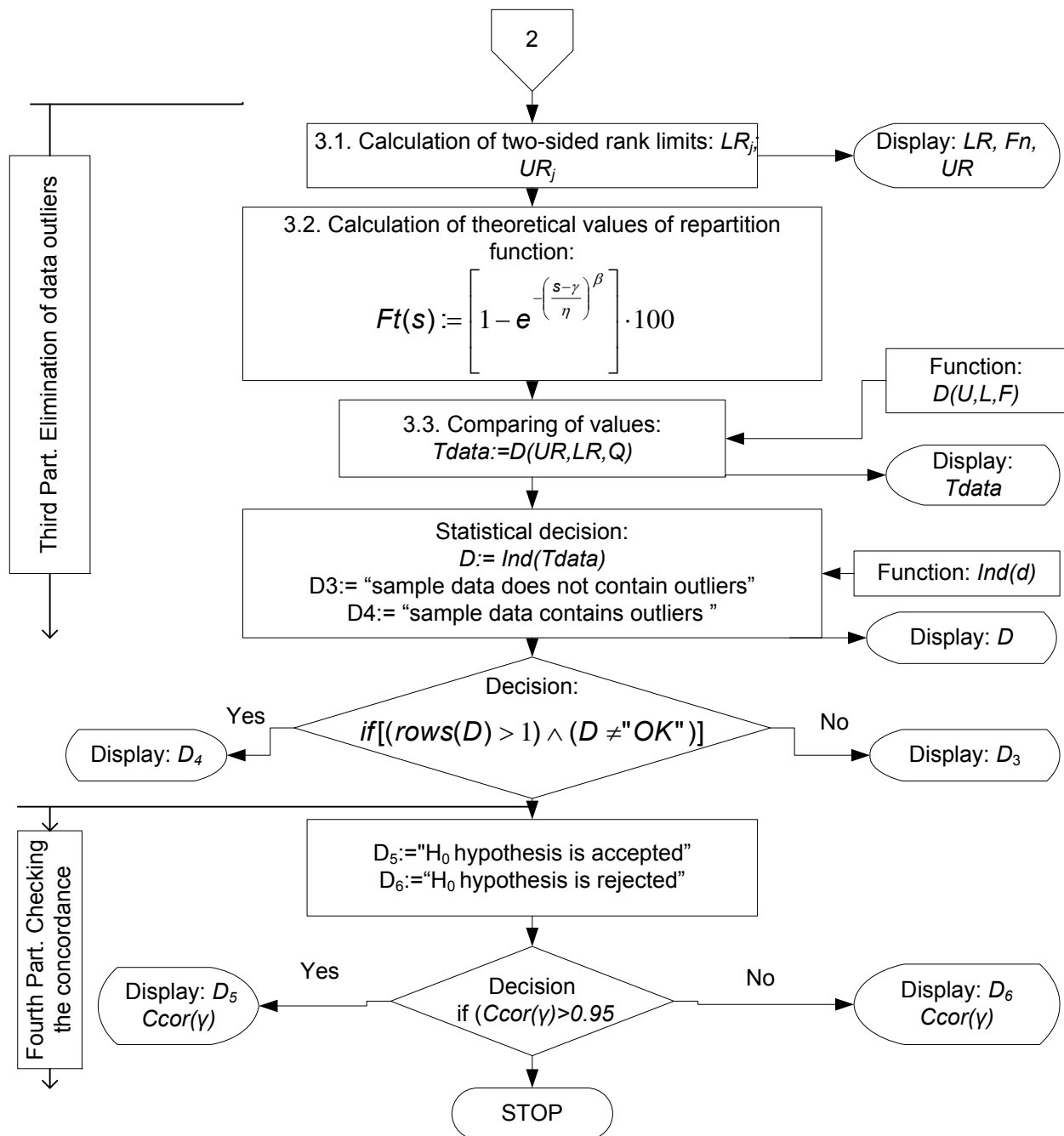


Fig.1. The logic scheme of the MathCAD program

#### 4. CASE STUDY

The data used within this case study were collected during a time of 1 year, respectively January - December 2004. During this period, the working of a *Slicer SM-H* - type production equipment specialized in processing wood into thin plates having thicknesses between 0,1- 0,2 mm has been followed. The equipment has worked in continuous running, 24 hours of 24, seven days from seven.

The equipment is provided by automate recorders of the staying times in repair. An informatic program written in Visual Basic for Application allowed to fastly determining of the times of good working. The following experimental data, in minutes, of the times of working without failure, have been obtained:

1400 9964 11440 3897 2782 7200 702 1426 500 9039 7167 22568

10027	2693	4310	2518	10080	4280	1355	6396	1144	1419	2854	2606
2673	10039	12955	8222	5382	2789	3895	7190	2301	2437	11339	5390
7189	5287	4187	2688	4308	2026	8084	4076	5728	4091	5326	2591
11513	2587	4278	4222	4206	1427	4247	2821	7101	11155	8502	2769
2405	5474	2840	4221	2537	2839	2786	4276	4212	1455	4246	8623
5747	2781	19684	7180	2750	5746	6869	4268	7128	14353	5502	5737
2863	18670										

For a correct reading of the observed data in the MathCAD program, their organization in a file ASCII named Date250408.txt has been necessary.

After the randomness of the experimental data has been checked, 46 iterations have resulted, the maximal lengths of iterations being of 7 successive numeric values, respectively 5746, 6869, 4268, 7128, 14353, 5502, 5737, higher than the sample's median, which is of 4246,5 minutes. These values are identified by the due position indexes, respectively: 77, 78, 79, 80, 81, 82, 83, and 84. The decision taken by the program, after having tested the randomness of experimental date is: *Decizion1 = "hypothesis  $H_0$ , concerning randomness is accepted"*.

Statistical modeling of the times of good working presumes also to check homogeneity of experimental data. By using the rank limits method, the program communicates the decision it has taken (fig.1): *"The sampling values contains outliers"*. It may be observed that the program displays the indexes of outliers: 31, 32, 48, 49, 83, 84, 85 for their identification and elimination.

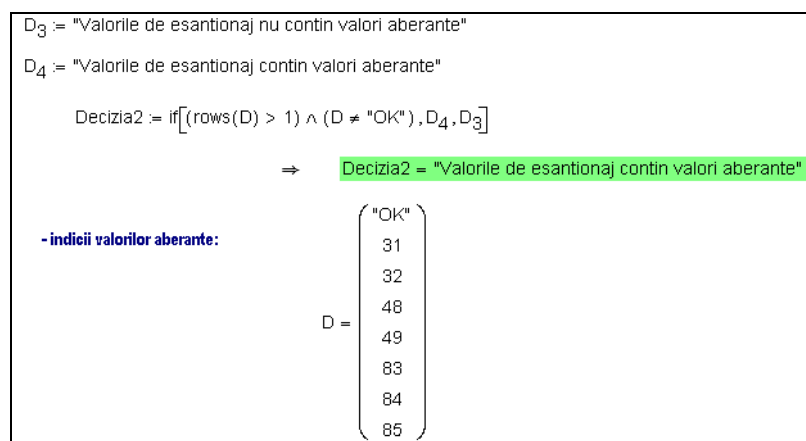
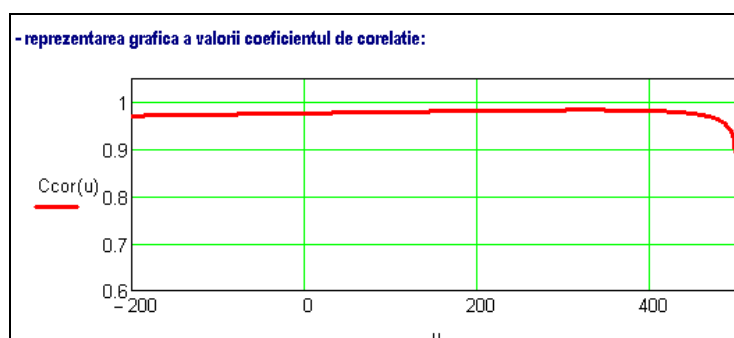


Fig. 2.– Identifying outliers – program sequence

In the fourth part, the program checks if experimental data are observing the hypothesis on the theoretical model adopted and it represents graphically (fig.2) the variation of the correlation coefficient according to the locating parameter of the Weibull distribution.



**Fig. 3. Variation of the correlation coefficient according to the locating parameter**

On the grounds of the calculated correlation parameter  $Ccor(\gamma) = 0.98384$ , the program displays (fig.3): *Decision 3 = "decision  $H_0$ , concerning the concordance of the sampling data is accepted"*.

```

4.1. Valoarea coeficientului de corelatie:

      Decizia3 := if(Ccor( $\gamma$ ) > 0.95, D5, D6)

4.2. Decizia statistica:

      Ccor( $\gamma$ ) = 0.98348

      ⇒ Decizia3 = "se accepta decizia  $H_0$ , privind concordanta datelor de esantionaj"

```

**Fig. 4. Value of the correlation coefficient– program sequence**

The calculated value of the correlation coefficient:  $Ccor(\gamma) = 0.98348$  confirms the hypothesis with regard to the good conformity between the Weibull distribution model and the times of working between accidental failures of the production equipment.

## 5. CONCLUSIONS

The MathCAD program presented hereby allows in a fast and easy way, to check the randomness of the samples represented by the times of working between failures of the production equipment, to eliminate outliers and to test the conformity with the Weibull distribution model, although the methods used are laborious.

The times of good working between accidental failures are correctly modeled by the three-parameter Weibull distribution.

## BIBLIOGRAPHY

1. Kececioglu, D. B. - *Reliability Engineering HandBook*. Vol. I și II, Inc. Englewood Cliffs, New-Jersey, 1993
2. Martinescu, I., Popescu, I.: *Fiabilitate*, Editura Gryfon, Braşov, 1995
3. Mărăscu-Klein, V., Toma, V.: *Managementul mentenanței*, Editura Universității Transilvania din Braşov, 2007
4. Monchy, F.: *Maintenance – Methodes et organisation*, Edition Dunod, 2003
5. Morariu, C.O.: *Optimizarea încercărilor de fiabilitate a rulmenților*, Teză de doctorat, Universitatea Transilvania din Braşov, 1999