THE USE OF TIN SOFTWARE IN STATISTICAL HYPOTHESIS TESTING. CASE STUDY

R. Drienovsky[1], M. Drienovsky[2], M. Boldea[3]

[1] Student of Banat University of Agricultural Sciences and Veterinary Medicine "Regele Mihai I al României" from Timisoara, Romania
[2]Mechanical engineer, Continental Automotive Group Romania
[3]Mathematics and Statistics, Banat University of Agricultural Sciences and Veterinary Medicine "Regele Mihai I al României" from Timisoara, Timișoara, Romania
[*]Corresponding author, e-mail: marius_boldea@usab-tm.ro

Abstract. *Hypothesis testing represents an important part of statistical analysis. Various types of software rely on this tool as their main part. The present paper introduces the software named TIN, which is designed by the authors, as well as the theoretical part that lays at the foundation of the programming code. In the case of a hypothetical example, we will use TIN for reading the experimental data. Then, still with the help of TIN, we will obtain the conclusions after applying the null hypothesis test, in two formats: .jpg and .txt. Their presentation is an easy one, as the main goal is to make TIN accessible to a large number of people who have little or no knowledge of statistics.*

Key words: *TIN, statistical analysis, software, null hypothesis test*

INTRODUCTION

Experimental practice (production) forces the researcher to compare two means from several measurements, i.e. two means of selection. This means the researcher has to establish whether the difference between them is real, reliable, or not. This difference might be unreliable (insignificant) due to the margin of error for each mean. Even if the means are equal, we cannot draw the conclusion that the two lots in which these were calculated are equal or even similar.

Therefore, in order to compare two means of selection, the so-called null hypothesis test is to be used.

MATERIAL AND METHOD

The mathematical foundation is represented by the null hypothesis test for comparing two means of selection. It is well-known that if we consider:

- Two means of selection:

$$\bar{X}_1 = \frac{\sum_{i=1}^{n_1} x_i}{n_1}, \qquad \bar{X}_2 = \frac{\sum_{i=1}^{n_2} y_i}{n_2}, \tag{1}$$

and the difference between them:

$$d = \bar{X}_1 - \bar{X}_2; \tag{2}$$

- the dispersion of the difference

$$D(d) = \frac{D_1}{n_1} + \frac{D_2}{n_2}, \tag{3}$$

then parameter t is calculated, which depends on the difference between the means and also on the dispersion of the difference $D(d)$, like this:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{D(d)}} \quad (4)$$

Three cases are distinguishable here:

a) Dispersions $D_1(X_1)$ and $D_2(X_2)$ are known, and $n_1, n_2 > 30$

In this case the distribution is normal and the value is calculated according to formulas (3) and (4).

b) Dispersions $D_1(X_1)$ and $D_2(X_2)$ are unknown, and $n_1, n_2 > 30$

In this case, variable t has also a normal distribution. Dispersions $D_1(X_1)$ and $D_2(X_2)$ are estimated by:

$$\tilde{D}_1 = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)}{n_1 - 1} \text{ and } \tilde{D}_2 = \frac{\sum_{i=1}^{n_2}(x_i - \bar{x}_2)}{n_2 - 1}, \quad (5)$$

and t is calculated with relations (5), (3) and (4):

c) Dispersions $D_1(X_1)$ and $D_2(X_2)$ are unknown, and $n_1, n_2 < 30$

In this case, variable t cannot be considered as having a normal distribution, the sampling fluctuations being large. It will have a new distribution called Student distribution with $\nu = n_1 + n_2 - 2$ degrees of freedom. Dispersions $D_1(X_1)$ and $D_2(X_2)$ are unknown but are considered equal and from (5) we have:

$$D = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1)}{n_1 - 1} = \frac{\sum_{i=1}^{n_2}(x_i - \bar{x}_2)}{n_2 - 1} = \frac{\sum_{i=1}^{n_1}(x_i - \bar{x}_1) + \sum_{i=1}^{n_2}(x_i - \bar{x}_2)}{n_1 + n_2 - 1}, \quad (6)$$

and (3) becomes:

$$D(d) = D\,\frac{n_1 + n_2}{n_1 n_2} \quad (7)$$

Taking into account (6) and (7) t in (4) becomes:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{D}} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}. \quad (8)$$

Once t is calculated with one of the cases above, it is compared with the theoretical interval:

$$(-t_0, t_0), \quad (9)$$

Resulted from the normal (Gauss) distribution or the Student distribution.

If a comparison is made between the means of coupled observations, we will consider two samples with the following measurements:

$$x_1, x_2, \dots, x_n \tag{10}$$

$$y_1, y_2, \dots, y_n \tag{11}$$

If there are good reasons for a pairing to be established between the two samples, then the null hypothesis test is changed in the following way:

the means are no longer calculated with relation (1) and the difference between means with relation (2), but the differences between pairs:

$$d_i = x_i - y_i \tag{12}$$

And after that, the mean between differences is calculated:

$$d = \frac{1}{n}\sum_{i=1}^{n} d_i \tag{13}$$

and the test continues.

RESULTS AND DISCUSSIONS

The programming was done with Blitz Plus free and it includes various modules assembled in the generator code of the TIN software. Figure 1 presents the selection of the folder in which raw data are stored:

```
;------- se selecteaza FOLDERUL
        numarul_folderului_selectat =SelectedGadgetItem(introducere_nume_1_box) ;Se aplica doar cand se intra la serch FOLDER
        If EventSource() =introducere_nume_1_box          ;<------------------------------------------------------------
            SetGadgetText(introducere_nume_1,GadgetItemText(introducere_nume_1_box,numarul_folderului_selectat) )

            ClearGadgetItems introducere_nume_2_box
            ;-----------------------
            curentul_folder$ =ReadDir(CurrentDir()+GadgetItemText(introducere_nume_1_box,numarul_folderului_selectat))
            ;AddGadgetItem(introducere_nume_2_box,"")
            Repeat
                nume_folder$ =NextFile(curentul_folder)
                If nume_folder <>".." And nume_folder <>"."
                    If FileType(nume_folder)<>2 Then AddGadgetItem(introducere_nume_2_box,nume_folder)
                    ;AddGadgetItem(introducere_nume_2_box,nume_folder)
                EndIf
                If nume_folder ="" Then Exit
            Forever
            ;-----------------------
            SetGadgetText(introducere_nume_2,GadgetItemText(introducere_nume_2_box,0) )
            SelectGadgetItem(introducere_nume_2_box,0)
            ;GadgetItemText(introducere_nume_1_box,0)
        EndIf
;-------
```

Figure.1 Opening and memorizing the content of the selected folder

The user-friendly work interface is the one below:

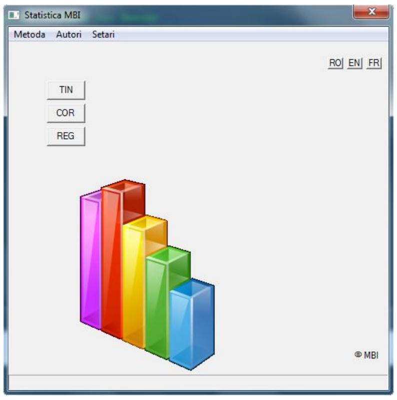

Figure 2 The menu window of TIN

We chose to give a hypothetical case as an example of a case study. We are given two wheat varieties A and B that were sown on similar fields for each pair. In order to determine the average yield per hectare we made 10 determinations, obtaining x_i *and* y_i as in the table below:

Variety *A* (x_i *kg/ha)*	980	1305	960	1140	1130	1080	1090	1270	1250	1100
Variety *B* (y_i *kg/ha)*	910	1245	910	1220	1010	990	1030	1170	1200	1060

We have to decide whether the differences between the average yields per hectare are significant.

After selecting the folder and the file in which raw data are stored, the software reads these data as in Figure 3:

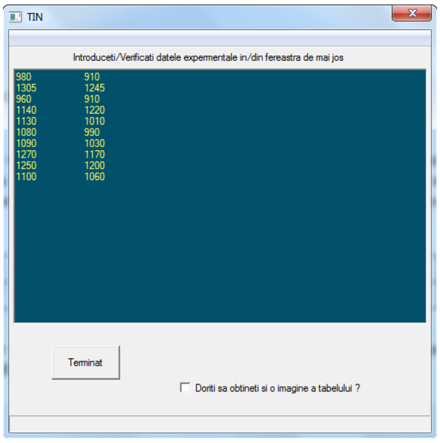

Fig. 3 Reading the data introduced

After the raw data are processed, TIN has the possibility to return the information in either a .txt file, or a .tiff image. Figures 4 and 5 present two print screens in which the processed data are presented as .tiff images.

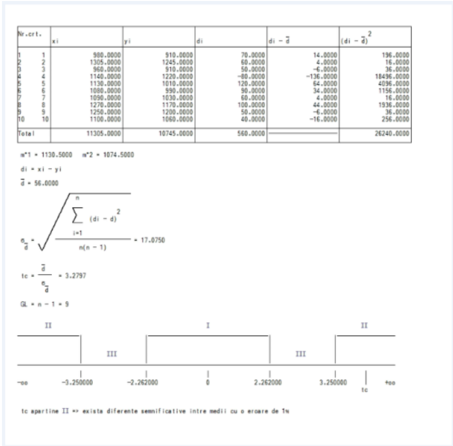

Figure 4 The result with coupled pairs

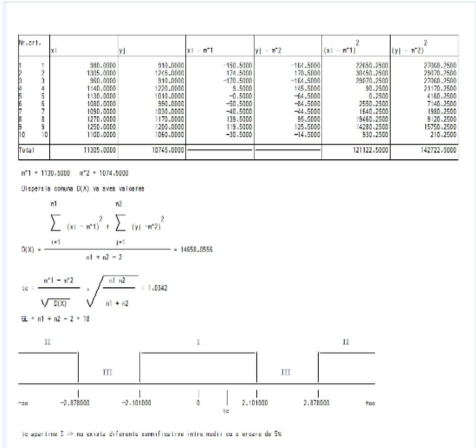

Figure 5 The result with uncoupled pairs

The example chose to illustrate this paper shows that it works both for couples frequencies and for uncoupled frequencies. The conclusion, in the case of uncoupled frequencies we will obtain the lack of significant differences, and in the case of coupled frequencies, we will obtain the existence of significant differences.

CONCLUSIONS

It is worth mentioning that coupled observations (where they are possible) is a finer method that can highlight a significant difference where the usual method cannot, when using the same set of data.

The use of TIN software is an easy one both for people who find it easy to work with experimental data (with good knowledge of statistics) and for farmers, in an attempt to ease and optimize their work.

BIBLIOGRAPHY

1.ANGHEL, C., M. BOLDEA. 2003. A new distribution law in statistics. Stochastic Analysis and Applications, Vol 3: 1-6.

2.ANGHEL, C., M. BOLDEA. 2007. A method for determining the production function. Differential Equations and Applications, Vol 4 4: 9-14.

3.BOLDEA, M. 2010, Probabilități și statistică matematică. Teorie și aplicații, Ed. Mirton, Timişoara.

4.BOLDEA, M., F. SALA, I. RADULOV, F. CRISTA. 2010. A Mathematical Model on the Dependence between the Agricultural Production and Chemical Fertilizers. Numerical Analysis and Applied Mathematics, Vols I-III 1281: 1363-1366. doi:10.1063/1.3497978.

5.SALA, F., G. ARSENE, O. IORDANESCU AND M. BOLDEA. 2015. Leaf area constant model in optimizing foliar area measurement in plants: A case study in apple tree. Scientia Horticulturae 193: 218-224. doi:10.1016/j.scienta.2015.07.008.

6.SALA, F., M. BOLDEA, H. RAWASHDEH AND I. NEMET. 2015. Mathematical Model for Determining the Optimal Doses of Mineral Fertilizers for Wheat Crops. Pakistan Journal of Agricultural Sciences 52: 609-617.

7.http://www.blitzmax.com/Home/_index_.php