ANALYSIS OF VARIANCE USED IN THE STUDY OF THE RELATIONSHIP BETWEEN VARIABLES

Associate Professor Ph.D. Elisabeta R. ROŞCA
„Stefan cel Mare" University of Suceava, Romania
elisabetar@seap.usv.ro

Abstract:

The paper "Analysis of Variance Used in the Study of the Relationship between Variables" presents an introduction in ANOVA, which has a different methodology depending on the number of factors taken into study. The paper emphasizes that ANOVA can be used in the case in which for the independent variable has used a simple group, but the dependent variable is presented under the form of variants in case of each group or when was used a combined group for both variables. The ANOVA involved also the used of Fischer -Snedecor test for testing the signification of group factors. The paper present the basic concepts and metho dology of ANOVA and the unifactorial and bifactorial methodology of analysis of variance.

Key words: analysis of variance (ANOVA), estimations of population variance, variance among sample means, variance within the samples, total, factorial or residual v ariances, one-factor or two factors analysis of variance.

INTRODUCTION

The analysis of variance (ANOVA) was introduced by the statistician R. A. Fisher and one of its use is for verification the measure in which the real values of the characteristic divert from the theoretical values, calculated as a rule as means and as regression equations, as well the measure in which these variances are depended on the group factor or not. The analysis of variance has at the base of group method, which allows the separation of the influence of the essential factors from the influence of random factors, on the resultant characteristic. [7], [9]

Depending on the number of factors (one, two or more) which has an influence about the variation of resultant variable there are unifactorial, bifactorial or multifactorial models of analysis of variance.

ANALYSIS OF VARIANCE – BASIC METHODOLOGY

The analysis of variance is a method which permits the study of the relationship between variables by testing the signification of difference between means of many samples. Using this method we can make inferences, if the samples are extracted from populations which have the same mean. In this context, in practice the analysis of variance can be used for the comparison the distance going through by consuming some different brands of gasoline, the testing of some didactic methods after their performances in teaching, the comparison of incomes obtain in the first year by the graduates of some faculties of same profile etc. In such kind of cases we will compare the means of some samples and the analysis of variance is used to decide if the samples were extract from population with equal means. So, if we write down the means of populations with: $\mu_1, \mu_2, \mu_3, \ldots,$ the hypothesis which must be tested are:

$H_0 : \mu_1 = \mu_1 = \mu_3 = \ldots \leftarrow$ the null hypothesis

$H_1 : \mu_1, \mu_2, \mu_3, \ldots$ are not all equal $\leftarrow$ the alternative hypothesis

If we consider, for example, the case of using the three didactic methods in a faculty, by testing we can draw the conclusion that between the samples means are not significant differences and so the option for a method will not determine the increase of the efficiency of teaching, but if we will find significant differences between samples means, too great to be determined by change the standard error, the conclusion is that the didactic method has an influence about the performances in teaching and as a result the method can be improved. [16], [17], [19].

For using the analysis of variance we must to consider the assumption that each sample extracted from a normal population and each of this population has the same variance, σ^2. If the sample volumes are enough great is not necessary the assumption of normality. [15], [16].

In the context of the previous presentation, the analysis of variance is based on the comparison of two different estimations of variance for total population, σ^2. One of these estimations can be computed using the variance between the sample means and the other can be determined from the observed values of samples. The two estimations of the variance of total population are compared and they must be about equal as value when the null hypothesis is true. If the null hypothesis is not true, then the two estimations will have considerable different values. In the application of analysis of variance are three steps, namely:

1. the determination of an estimation of variance of population on the basis of variance between the sample means;
2. the determination of an estimation of variance of population on the basis of observed sample values;
3. the comparison of the two estimations. If they are equal in value the null hypothesis is accepted.

So the first step supposes the computing the variance from the sample means. If we compute the sample variance with the formula:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} \tag{1}$$

where:

n represents the number of units from the sample.

If we consider k samples, which have the means wrote down $\bar{x}$ and the mean of samples means wrote down $\bar{\bar{x}}$, the variance between the sample means is computed with the formula:

$$S_{\bar{x}}^2 = \frac{\sum (\bar{x} - \bar{\bar{x}})^2}{k-1} \tag{2}$$

The standard error of mean is defined as standard deviation of all possible samples of a given volume and it is computed with the formula:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \tag{3}$$

where:

σ represents the standard deviation of total population.

If we square up this formula then the variance of population, σ^2 is determined with the formula:

$$\sigma^2 = \sigma_{\bar{x}}^2 \cdot n \tag{4}$$

where:

$\sigma_{\bar{x}}^2$ represents the square of standard error (the variance between sample means - $S_{\bar{x}}^2$).

The indicator $S_{\bar{x}}^2$ can be calculated from the observed values in samples and can replace the value $\sigma_{\bar{x}}^2$ in the equation (4), which gives the possibility to estimate the variance of population using the formula:

$$\sigma_{\bar{x}}^2 = S_{\bar{x}}^2 \cdot n = \frac{\sum n\left(\bar{x} - \bar{\bar{x}}\right)^2}{k-1} \quad (5)$$

In the case in which are used the samples with different volume, noted n_j, the estimation of variance for population on the basis of variance between samples means will make with the formula:

$$\hat{\sigma}^2 = \frac{\sum n_j \left(x - \bar{x}\right)^2}{k-1} \quad (6)$$

As we can observe, the second step in the application of analysis of variance consist of determination an estimation for the variance of population on the base of observed values of samples.

The variance inside the samples can be calculated with the formula:

$$S^2 = \frac{\sum \left(x - \bar{x}\right)^2}{n-1} \quad (7)$$

If we assume the premise already established that the population from which are extract the samples have the same variance we can use any variance of sample as the second e stimation of variance of population. In a statistical way, we can obtain a better estimation of the variance of population by using the weight mean of sample variance. In this case, the general formula for computing of the second estimation of the variance of population based on the variance inside the samples is:

$$\hat{\sigma}^2 = \sum \left(\frac{n_j - 1}{n_j - k}\right) \cdot s_j^2 \quad (8)$$

where:

n_j represents the dimension of sample j;

s_j^2 represents the variance of sample j;

k represents the number of samples;

$n_T = \sum n_j$ represents the total sample size.

Such a formula for estimation has the advantage that it used all the information we have, not only a part of these.

The third step in the application the analysis of variance consist of the comparison the two estimations of variance of population by computing their F ratio using the formula:

$$F = \frac{First\_estimate\_of\_the\_population\_variance\_based\ on\_the\_variance\_among\_the\_sample\_means}{Second\_estimate\_of\_the\_population\_variance\ based\_on\_the\_variances\_within\_the\_samples} \quad (9)$$

The numerator and the denominator will be equal if the null hypothesis is true and as much the value of F ratio is close by 1 such we are inclined to accept the null hypothesis. If the F ratio becomes great we are inclined to reject the null hypothesis and to accept the alternative hypothesis. The logic meaning of statistic test F is that *for different populations the estimated variance on the basis of sample means tend to be greater than the estimated variance on the basis of variances inside the samples and the value of F ratio tend to increase. This brings us to the rejecting the null hypothesis.* [16], [17], [19] In the context of general methodology of analysis of variance is important to analyse if the statistic test F has a particular sample distribution, if the null hypothesis is true. In reality, the F distribution is a whole family of distributions, each of them being identif ied by a pair of freedom degrees, different from the distributions t and χ^2, which have a single value for the number of freedom degrees. The first number from the pair of freedom degrees is the numerator of F ratio and the second number is the denominator.

The F distribution is one single mode distribution. The characteristic of the curve of distribution F depend on the number of freedom degrees f rom the numerator and the denominator of F ratio. But generally the F distribution is skewed to the right and tends to become more symmetrical as the number of freedom degrees is increased.

Also, another problem is the establishing of the variance of population on the basis of variance between sample means, for which is required the computing of the expression $\sum\left(\bar{x}-\bar{\bar{x}}\right)^2$, the number of terms of this being equal with the number of samples. So, the num ber of freedom degrees for the numerator of F ratio is always smaller with 1 face to the number of sample that is:

$$number\_of\_freedom\_degrees\_in\_numerator\_of\_the\_F\_ratio = \left(numer\_of\_samples - 1\right) \quad (10)$$

For the computing the variance inside the samples are used all the samples. So, if we have j samples, we use n_j values $\left(x-\bar{x}\right)$ for the computing the sum $\sum\left(x-\bar{x}\right)$ for each sample. The number of freedom degrees for the numerator of F ratio is computing with the formula:

$$number\_of\_freedom\_degrees\_in\_denominator\_of\_the\_F\_ratio = \sum\left(n_j - 1\right) = n_T - k \quad (11)$$

where:
n_j represents the volum of sample j;
k reprezents the number of samples;
$n_T = \sum n_j$ represents the total sample size.

For the application the test F we must have a table of distribution F in which the colons represent the number of freedom degrees for numerator and the lines represent the number of freedom degrees for the denominator. For each level of significance there are different tables. So, according to a certain freedom degrees and a certain level of significance is selected the value of F ratio from the table and is compared wi th the computed value. If $F_{calculat} \succ F_{tabelar}$ then the null hypothesis is rejected and if not then the null hypothesis is accepted. [15], [16], [17]

The methodology of analysis of variance must be applied with much discrimination because for it has the ability to be on the base of some significant decisions it must to exist the certainty that all the factors are effective controlled. The presented methodology is referred at one single factor which has an influence about a process (we used, for example, the case of using of more didactic methods with different efficiency) but the method of analysis of variance can be used also for study the influence of more factors about a process. So, we continue with the unifactorial and bifactorial analysis of variance. [14], [20]

THE UNIFACTORIAL ANALYSIS OF VARIANCE

Unifactorial analysis of variance has at the base the group depending on the factor X_i. Starting from the hypothesis that the conditioned means by the group factor, that is $\bar{y}/x_i = \bar{y}_i$, represents the typical values which are forming at the level of each group and the general mean $\bar{y}$ is the typical value for the whole population, it can be formulated two conclusions such as:

a) the variation of resultant variable Y is dependent by the group factor (x_i);

b) the variation of resultant variable Y is independent by the group factor (x_i).

To specify which conclusion is real is necessary to take into consideration the measure in which the individual values y_i are diverted from the conditioned means by the group factor and from the general mean $(\bar{y})$. These deviations are synthetic reflections of the result of the way of association of factors which determine the variation of variable Y. [1], [5], [10]

The basic idea of analysis of variance is that to arrived at one of previous formulated conclusions is necessary to decompose the sum of squares of the deviations from the general mean $\sum_{j=1}^{m}(y_j-\bar{y})^2 n\cdot n_{.j}=\left(\sum_i\sum_j(y_j-\bar{y}_i)^2 n_{ij}\right)$ into a number of components, where each part corresponds to a real or suppositional source of variation of means.

For this we consider that the resultant variable Y with the individual values distributed on the groups $y_1, y_2, ..., y_{n_i}$ is influenced by the independent variable X_i, which has the individual values $x_1, x_2, ..., x_r$. The observed data was distributed on r groups, where each group contains also the individual values of variable Y, about that in the model exist the hypothesis that it is a normal distribution. So we obtain the conditioned distribution from the Table no 1 .

Table no 1. The conditioned distribution of variable Y

The groups of population by the independent characteristic X	**Values of result characteristic Y**	**Number of individual values of characteristic Y**	**Group means (conditioned by x_i)** $y/x_i=y_i$
0	1	2	3
x_1	$y_{11}y_{12}\ldots y_{1j}\ldots y_{1n}$	n_1	$\bar{y}_1$
x_2	$y_{21}y_{22}\ldots y_{2j}\ldots y_{2n}$	n_2	$\bar{y}_2$
$\vdots$	$\vdots\vdots \quad \vdots \quad \vdots$	$\vdots$	$\vdots$
x_i	$y_{i1}y_{i2}\ldots y_{ij}\ldots y_{in}$	n_i	$\bar{y}_i$
$\vdots$	$\vdots\vdots \quad \vdots \quad \vdots$	$\vdots$	$\vdots$
x_r	$y_{r1}y_{r2}\ldots y_{rj}\ldots y_{rn}$	n_r	$\bar{y}_r$

Note: $1\le i\le r; 1\le j\le n_i$

On the base of conditioned distribution of variable Y, it can compute:

- the mean of values y_{ij} from each group i (the mean of Y conditioned by X_i), using the formula:

$$\bar{y}_i=\frac{\sum_{j=1}^{n_i}y_{ij}}{n_i},\ i=1\div r \tag{12}$$

- the general mean of values y_{ij} using the formula:

$$\bar{y}=\frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i}y_{ij}}{\sum_{i=1}^{r}n_i}=\frac{\sum_{i=1}^{r}\bar{y}_i n_i}{\sum_{i=1}^{r}n_i}=\frac{\sum_{j=1}^{n}y_i}{n} \tag{13}$$

where:

$\sum_{i=1}^{r} n_i = n$ which means the number of units from population (or from the random selected sample) distributed on r groups.

The measure in which the individual values y_{ij} are diverted from $\bar{y}$ as a result of the action of whole factors of influence can be written so:

$$\Delta_y^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left[\left(y_{ij} - \bar{y}_i\right) + \left(\bar{y}_i - \bar{y}\right)\right]^2 =$$

$$= \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(\bar{y}_i - \bar{y}\right)^2 + 2\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)\left(\bar{y}_i - \bar{y}\right) =$$

$$= \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2}_{S_2} + \underbrace{\sum_{i=1}^{r}\left(\bar{y}_i - \bar{y}\right)^2 n_i}_{S_1} + \underbrace{\sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)\left(y_i - \bar{y}\right)}_{\operatorname{cov}(x_i, z_i)} \qquad (14)$$

where:

$$\Delta_y^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}\right)^2 \qquad (15)$$

represents the total variance which emphasized the influence of essential factor X and unessential factors Z_i about the variable Y;

$$\Delta_{y/x}^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(\bar{y}_i - \bar{y}\right)^2 = \sum_{i=1}^{r}\left(\bar{y}_i - \bar{y}\right)^2 n_i \qquad (16)$$

represents the factorial variance (systematic) between groups, which emphasized the influence of variation of group factor X (the factor with systematic action) about the resultant variable Y;

$$\Delta_{y/z}^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}\left(y_{ij} - \bar{y}_i\right)^2 \qquad (17)$$

represents the residual variance inside the groups, which emphasized the influence of unessential, residual factors (included in Z_i) which operate inside of each r groups;

$\operatorname{cov}(x_i, z_i)$ represents covariance between the essential group factor X and the rezidual factors (Z_i).

In the hypothesis in which $\operatorname{cov}(x_i, z_i) = 0$, then:

$$\Delta_y^2 = \Delta_{y/x}^2 + \Delta_{y/z}^2 \qquad (18)$$

Considering:

- $\Delta_{y/x}^2 = S_1^2$ which represents the sum of squares of the deviations of group means from the general mean weighted with the frequencies of groups;
- $\Delta_{y/z}^2 = S_2^2$ which represents the sum of squares of the deviations between the observed values in each group i and its group mean in the general population;
- $\Delta_{y/x}^2$ şi $\Delta_{y/z}^2$ are independent values;

$$S_y^2 = S_1^2 + S_2^2 \qquad (19)$$

The three variances S_y^2, S_1^2 and S_2^2 are square forms of the deviations of variants y_{ij}. It can be proved that S_y^2 can became a sum of squares $\sum_{j=1}^{n-1} y_j^2$ with the rank not bigger than $(n-1)$.

S_1^2 is a sum of r liniar forms with the rank not bigger than $(r-1)$ and S_2^2 is the sum of n liniar forms which fulfil r independent relations, what allows to obtain a rank not bigger $(n-r)$. So, the rank (the number of freedom degrees) of total variance is the sum of variances S_1^2 and S_2^2, that is:

$$n-1=(r-1)+(n-r) \quad (20)$$

The rank or the number of freedom degrees emphasizes the number of independent elements necessary to define an ensemble. The number of freedom degrees is obtained, generally, by subtraction from the number of elements considered simultaneous the number of conditioned means established for the population (for example, if for the calculus of total variation is taken into consideration only the general mean, for the residual variance are taken into consideration r group means). [1], [4], [6]

By reporting the three variance (S_y^2, S_1^2, S_2^2) to the number of freedom degrees proper with each of them are obtained corrected variances (s_i^2) or estimations of variances (the general variance, the variance between groups and the variance inside the group). On, to verify the signification of group factor is computing the ratio between the corrected dispersion between groups (the factorial variance) and the corrected dispersion inside of groups (the residual variance).

Table no 2. The model of unifactorial analysis of variance

The type of variance	**The variation (the sum of squares of deviations)**	**The number of freedom degrees**	**The estimations of variances (the corrected variances)**	$F_{computed}$
A	1	2	3	4
The factorial variance (systematic or between groups)	$S_1^2=\sum_{i=1}^{r}(\bar{y}_i-\bar{y})^2 n_i$	$r-1$	$s_1^2=\frac{S_1^2}{r-1}$	$F=\frac{s_1^2}{s_2^2}$
The residual variance (inside the groups)	$S_2^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2$	$n-r$	$s_2^2=\frac{S_2^2}{n-r}$	-
The total variance	$S^2=\sum_{i=1}^{r}\sum_{j=1}^{n_i}(y_{ij}-\bar{y})^2$	$n-1$	$s^2=\frac{S^2}{n-1}$	-

From presented elements result that the scheme of calculus for the model of unifactorial analysis of variance with the date systematized on the base of simple group can be synthetic presented as in the Table no. 2. For a level of signification α, selected according to the number of freedom degrees $r-1$ şi $n-r$, is searching in the table of distribution Fischer-Snedecor, the value $F_{\alpha;r-1;n-r}$ named tabelar value. [3], [9]

The interpretation is made like that:

- if between the computed value and value taken from the table exist the relation: $F_{r-1;n-r}$ > $F_{\alpha;r-1;n-r}$, than the hypothesis that the resultant variable Y is significant dependent by the group factor X is accepted;

- if between the computed value and the value taken from the table exist the relation $F_{r-1;n-r} < F_{\alpha;r-1;n-r}$, than the hypothesis that the resultant variable Y is independent from X is accepted.

Generally this type of model for analysis of variance is possible to be applied when for the individual values of variable X was used a simple group of data and for the variable Y the individual values were centralized conditioned by the group from which they ar e part, as distinct individual values, the models being possible to be applied by two or more characteristics for which were elaborated bifactorial or multifactorial models of analysis of variance. Such kinds of models must emphasize the separate influence of each factor about the factorial characteristic and the relationship between factors. [2], [8], [11]

THE BIFACTORIAL ANALYSIS OF VARIANCE

In the model of bifactorial analysis of variance the observed data are group by two variables. So, it is considered an experiment in which is followed the comparison of the effects of r different treatments. The total number of experiments is devided in p blocks of equal volums. Each block is devided in r equal objects of experience about will be applied the r different treatments. The value y_{ij} obtained at the object of experience about which was applied the treatment i and which is included in the block j is noted y_{ij}. At the same time is supposed that the variants y_{ij} are independent and they have a normal distributi on. [13], [18]

From the obtained data it can be computed arithmetic means, using the formula:

$$\bar{y}_i = \frac{1}{p}\sum_{j=1}^{p} y_{ij} \qquad \text{the mean for treatment } i; \qquad (21)$$

$$\bar{y}_i = \frac{1}{r}\sum_{i=1}^{r} y_{ij} \qquad \text{the mean for block } j; \qquad (22)$$

$$\bar{y} = \frac{1}{rp}\sum_{ij} y_{ij} \qquad \text{the general mean of selection.} \qquad (23)$$

The total variation is decomposed in the variation determined by each of these two group factors and the variation determined by the re sidual factors. So, is valid the id entity:

$$\sum_{ij}\left(y_{ij} - \bar{y}\right)^2 = p\sum_{i}\left(\bar{y}_i - \bar{y}\right)^2 + r\sum_{j}\left(\bar{y}_j - \bar{y}\right)^2 + \sum_{ij}\left(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}\right)^2 \qquad (24)$$

or

$$S_T = S_1 + S_2 + S_3 \qquad (25)$$

where: $S_T = \sum_{ij}\left(y_{ij} - \bar{y}\right)^2$; $S_1 = p\sum_{i}\left(\bar{y}_i - \bar{y}\right)^2$; $S_2 = r\sum_{j}\left(\bar{y}_j - \bar{y}\right)^2$; $S_3 = \sum_{ij}\left(y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y}\right)^2$

The ranks of variances S_T, S_1, S_2, S_3 are respective: $rp-1$; $r-1$; $p-1$; $(r-1)(p-1)$, from where results the equality:

$$rp - 1 = (r-1) + (p-1) + (r-1)(p-1) \qquad (26)$$

By dividing the variance at the number of freedom degrees we obtain the following undisplaced dispersions: s^2, the general dispersion; s_1^2 the dispersion for the factor „treatment"; s_2^2 the dispersion for the factor blocks; s_3^2 the residual dispersion.

The interpretation of the results can be made using the test F, respective:

$$F_{(r-1);(r-1)(p-1)} = \frac{s_1^2}{s_2^2} \text{ is compared with } F_{q,(r-1);(r-1)(p-1)}$$

and

$$F_{(p-1);(r-1)(p-1)} = \frac{s_2^2}{s_3^2} \text{ is compared with } F_{q;(p-1);(r-1)(p-1)}.$$

The possible conclusions are that a significant influence can have the both group factors ("treatments" and respective "blocks"), one single group factor or any of these factors to have a significant influence.

By generalization, we can say that in the case of bifactorial analysis of variance the registered data are grouped by two variables *A* and *B*. So, are obtained *r* groups after one variable, *p* groups after other variable and *rp* subgroups. [12], [18]

The scheme of bifactorial analysis of variance is presented in Table no. 3.

Table no 3. The model of bifactorial analysis of variance

The kind of variance	**The sum of square deviation (the variance)**	**The number of freedom degrees**	**The corrected variance**	***F* calculated**
A	1	2	3	4
Factorul A	$S_1 = p\sum_j (\bar{y}_i - \bar{y})^2$	$r-1$	s_1^2	s_1^2 / s_3^2
Factorul B	$S_2 = r\sum_j (\bar{y}_j - \bar{y})^2$	$p-1$	s_2^2	s_2^2 / s_3^2
Reziduală	$S_3 = \sum_{ij} (y_{ij} - \bar{y}_i - \bar{y}_j + \bar{y})^2$	$(r-1)(p-1)$	s^2	1
Total	$S_T = \sum_{ij} (y_{ij} - \bar{y})^2$	$rp-1$	s^2	-

In the practice are frequent the cases in which the data are presented under the form of groups, case in which intervene the frequencies and the calculus formulae are proper adapted to these cases.

CONCLUSIONS

Near by the utility in the verification of the form and the degree of interdependence between variables, the analysis of variance is length used in the verification of the signification of the group factor, in the programming of the experiments with the aim to improve the performances of an industrial, agricultural processes etc., the model having an increasing complexity as the group becomes more complex.

BIBLIOGRAPHY:

[1] Andrei, T., Stancu, S., Pele, D. T., *Statistica. Teorie şi aplicaţii*, Editura Economică, Bucureşti, 2002

[2] Anghelache, C., *Statistică generală,* Editura Economică, Bucureşti, 1999

[3] Anghelache, C., *Statistică. Teorie şi aplicaţii*, Editura Economică, Bucureşti, 1998

[4] Baron, T., Anghelache, C., Ţiţan, E., *Statistică*, Editura Economică, Bucureşti, 1996

[5] Baron, T., Biji, E. M., Tövissi, L., et al., *Statistică teoretică şi economică*, Editura Didactică şi Pedagogică – R.A., Bucureşti, 1997

[6] Baron, T., Ţiţan, E., Matache, S., Ciuchiţă, L., *Manual practic de statistică*, Editura Expert, Bucureşti, 1999

[7] Bădiţă, M., Baron, T., Korka, M., *Curs de statistică*, Editura Sylvi, Bucureşti, 1993

[8] Bădiţă, M., Baron, T., Korka, M., *Statistica pentru afaceri*, Editura Eficient, Bucureşti, 1998

[9] Biji, M., Biji, E. M., Lilea, E., Anghelache, C., *Tratat de statistică aplicată*, Editura Economică, Bucureşti, 2002

[10] Biji, E. M., Lilea, E., Roşca, R. E., Vătui, M., *Statistică aplicată în economie*, Editura Universal Dalsi, Bucureşti, 2000

[11] Biji, E. M., Lilea, E. Voineagu, V., *Statistică*, Universitatea Creştină „Dimitrie Cantemir" Bucureşti, 1994

[12] Isaic-Maniu, Al., Mitruţ, C., Voineagu, V., *Statistică*, Editura Universitară, Bucureşti, 2004

[13] Isaic-Maniu, Al., Mitruţ, C., Voineagu, V., *Statistica pentru managementul afacerilor*, Editura Economică, Bucureşti, 1999

[14] Jaba, E., *Statistică*, Editura Economică, Bucureşti, 1998

[15] Keller, G., Warrack, B., Bartel, H., *Statistics for Management and Economics. A Systematic Approach*, Wadaworth Publishing Company, Belmont, California, 1988

[16] Levin, I., R., *Statistics for management*, Fourth Edition, Published by Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987

[17] Loftus, G. R., Loftus, F. E., *Essence of Statistics*, Second Edition, Published by Alfred A. Knopf, Inc., New York, 1988

[18] Roşca, R. E., *Aplicarea analizei varianţei în studiul dependenţei modificării unui fenomen sau proces de condiţiile în care se desfăşoară. Partea I -Metodologia de bază de analiză a varianţei*, în volumul Conferinţei cu participare internaţională "Tehnomus", Universitatea „Ştefan cel Mare" Suceava, 1997

[19] Roşca, R. E., *Utilizarea metodei ANOVA în cercetarea statistică*, în volumul Sesiunii ştiinţifice internaţionale „Economia românească-prezent şi perspective", Editura Universităţii Suceava, 2002

[20] Ţarcă, M., *Tratat de statistică aplicată*, Editura Didactică şi Pedagogică R. A., Bucureşti, 1998