VARIANCE ANALYSIS APPLIED TO BALNEARY TOURISM ON ROMANIAN SEASIDE

Mirela Secară and
Daniela Simona Nenciu[*]

Abstract:

By means of ANOVA variance analysis we have studied the size and frequency of deviation of the real values of statistic characteristic from the calculated theoretical values, as well as the level of dependence or independence of these variations of the group factor. Variance analysis applies especially when data are generated by selective research.

Statistic poll that I personally conducted on Romanian seaside balneary tourism has mainly investigated the attitude, opinions and motivations of balneary tourism demand on Romanian Black Sea coast. One of the objectives of this partial statistic research on seaside hydro tourism is represented by the duration of a balneary vacation that is influenced by the age of tourists as well.

In order to evince the relation between the two variables above we have processed the data of the sampling, we have sorted and grouped tourists by their age and duration of vacation and then we have applied the variance analysis to verify and demonstrate that balneary tourists' age induces the duration of a balneary vacation.

Keywords: variance analysis, statistic poll, balneary tourism, dispersion, coefficient of determination.

JEL Classification: C10, L80, L83

Introduction

To perform statistic characterization of the phenomena and mass social-economic processes to evince their specific governing laws, individual data obtained of methods of total or partial record are to be processed. Statistic polls are methods of partially recorded data. The domain of tourism has certain features whose measuring can only be achieved by means of selective research, or features that need the results of such research for a complete representation.

Opinion poll is the most used method of selective research performed on a group of people. In the field of tourism, this method can provide relevant data concerning particularities of the demand for tourism during a certain period when a poll is conducted. The essential feature of the opinion poll is the use of scientific

[*] Mirela Secară is Associate Professor of statistics at the Dimitrie Cantemir University in Bucharest. E-mail: mirela_secara@yahoo.com
Daniela Nenciu is Associate Professor of marketing and tourism at the Dimitrie Cantemir University in Bucharest. E-mail: nenciu75@yahoo.com

method that is different from the journalistic approach, for instance, or other methods that do not consider statistic values or reality as a whole.

Opinion survey provides important information concerning particularities of the demand for balneary tourism on Romanian seaside. The date when the survey was conducted evinces its main advantage as well, that is the information was accurately obtained, in short time and with relatively low expense.

Theoretical background for variance analysis

The investigation used to complete the opinion poll within this area of balneary tourism was conducted on a representative part of statistic number of tourists and it focused on the characteristics of the demand for balneary tourism on Romanian seaside, on tourists' behavior, opinions and needs.

Coordination of an opinion survey requires a set of objectives and their concise presentation with relevance for the set target. The clear definition of the objectives is important not only to the next stages, but also to the other phases of research.

One of the objectives of the survey statistic research on seaside balneary tourism is the duration of vacation and the manner that tourists' age may influence the duration of vacation in seaside spas.

In order to evince the dependence between the two variables above-mentioned, we have processed the data of the survey, then grouped the tourists according to age and duration of vacation and applied variance analysis to verify if the age of the tourists that came on Romanian seaside for balneary tourism may influence the duration of their vacation.

By means of variance analysis also known within scientific studies as ANOVA, the size and frequency of deviation of statistic characteristics from calculated theoretical values are being studied, along with the level of dependence or independence of these variations from the group factor. [Elena Maria Biji et al.(2002)].

In this way, the characteristic whose variation is being studied is considered as resultant variable, marked with y, and the characteristics that lead to data group are considered as independent or factorial variables, marked with $x_1, x_2, \ldots x_n$. [Elena Maria Biji et al.(2002)].

In case of a single group factor, the issue is to establish how much of the total variation is generated by the group factor and to what extent it is generated by other unregistered factors, also known as residual factors, considered as having constant action within all studied cases. Thus, total variation can be divided in two parts. [Elena Maria Biji et al.(2002)]:

- variation generated under the influence of x factor;
- variation generated by unregistered factors, known as residual factors z.

Variance analysis is based on group method that divides the observed units into groups according to the variance of group characteristic that is considered a factor of influence and then, the dispersion of terms within each group is determined and also among the formed groups. [Elena Maria Biji(2002)].

In order to perform variance analysis the following indicators are being used as: deviation of terms from their average (Δ^2), the degree of factor determination (η^2), and ratio coefficient among dispersions (F). In case of two-entrance table, when data are grouped and combined according to both *x* and *y* characteristics, the calculation formulas of the indicators are such as:

- total deviation is calculated according to marginal distribution of *y* characteristic, by means of formula: $\Delta_y^2 = \sum_{j=1}^{m} \left(y_j - \overline{y_0}\right)^2 n_j$; (1)

- deviation generated by x factor is calculated according to distribution of conditioned elements of y characteristic, by means of formula:

$$\Delta_{y/x}^2 = \sum_{i=1}^{r} \left(\overline{y_i} - \overline{y_0}\right)^2 n_i ; \quad (2)$$

- residual deviance**,** that is the deviance generated by unregistered factors that acts only inside the groups, is calculated according to conditioned distributions of *y* characteristic, summing the square deviations of all formed groups according to group factor variation that is as: $\Delta_{y/z}^2 = \sum_{i=1}^{r}\sum_{j=1}^{m} \left(y_j - \overline{y_i}\right)^2 n_{ij}$. (3)

There is the same relationship, among the three deviations meaning:

$$\Delta_y^2 = \Delta_{y/x}^2 + \Delta_{y/z}^2 . \quad (4)$$

Variance analysis applied within statistic research of seaside balneary tourism

As a result of procession and centralization of results of opinion poll conducted within the field of balneary tourism on Romanian seaside, tourists' distribution according to duration of vacation (question no. 4 of the questionnaire) and age is presented in Table no. 1:

Tourists' distribution according to duration of vacation and age Table no.1

Age group	Duration of vacation							
	18 days	15 days	12 days	10 days	8 days	7 days	6 days	Total
Up to 25	4	0	0	2	1	10	7	24
26-35	2	0	0	10	2	16	5	35
36-45	9	3	0	12	0	7	15	46
46-55	28	2	0	4	0	4	3	41
56-65	28	1	0	1	0	2	0	32
over 65	27	0	1	1	1	1	0	31
Total	**98**	**6**	**1**	**30**	**4**	**40**	**30**	**209**

Source: personally processed data

I shall verify by means of variance analysis if the age of the tourists that arrived on Romanian seaside for balneary tourism is a determinant factor for the duration of vacation. Data are presented in a table with two way entrance, being the result of combined grouping according to both characteristics (x and y), where frequencies are common and independent. Group factor – *x* variable – is represented by the age groups and the result variable *y* is the duration of vacation.

Null assumption indicates absence of influence of balneary tourists' age on the duration of vacation and alternative assumption indicates existence of connection between the two variables (dependence). Indicators that are necessary to the variance analysis are calculated according to data in Table no.2:

Tourists' distribution according to duration of vacation and age Table no.2

Tourist groups of *x* group characteristic and age	Duration of vacation (result variable y) – number of days								
	18	15	12	10	8	7	6	Partial frequencies	Partial averages of *y* variable
Up to 25	4	0	0	2	1	10	7	24	$\overline{y_1}$ =8,83
26-35	2	0	0	10	2	16	5	35	$\overline{y_2}$ =8,4
36-45	9	3	0	12	0	7	15	46	$\overline{y_{13}}$ =10,13
46-55	28	2	0	4	0	4	3	41	$\overline{y_4}$ = 15,12
56-65	28	1	0	1	0	2	0	32	$\overline{y_5}$ = 16,97
Over 65	27	0	1	1	1	1	0	31	$\overline{y_6}$ = 16,87
Frequencies of total *y* variable	98	6	1	30	4	40	30	209	$\overline{y_0}$ = 12,72

Source: personally processed data

Calculated indicators are as:

-total deviation $\Delta_y^2 = \sum_{j=1}^{m} \left(y_j - \overline{y_0}\right)^2 n_j = 5738.34;$ (5)

-deviation generated by x factor $\Delta_{y/x}^2 = \sum_{i=1}^{r} \left(\overline{y_i} - \overline{y_0}\right)^2 n_i = 2672.25;$ (6)

-residual deviance $\Delta_{y/z}^2 = \sum_{i=1}^{r}\sum_{j=1}^{m} \left(y_j - \overline{y_i}\right)^2 n_{ij} = 3066.09.$ (7)

As we pointed out in the above, there is the same relationship among the three deviances, that is:

$$\Delta_y^2 = \Delta_{y/x}^2 + \Delta_{y/z}^2 = 2672.25+3066.09=5738.34. \quad (8)$$

According to obtained results, the coefficient of determination is as:

$$\eta_x^2 = \frac{\Delta_{y/x}^2}{\Delta_y^2} \cdot 100 = \frac{2672,25}{5738,34} \cdot 100 = 46,57\% \cong 47\%. \quad (9)$$

Thus, we may say that, as for the studied sampling, almost half (47%) of total variation is generated by the variation of the group factor (tourists' age) and the rest of it is the relative influence of the other factors. In order to establish the significance of the group factor, we shall have the ratio between the two dispersions. The two adjusted dispersions are as:

$$s_{y/x}^2 = \frac{\Delta_{y/x}^2}{r-1} = \frac{2672,25}{6-1} = 534,45 \quad (10)$$

and

$$s_{y/x}^2 = \frac{\Delta_{y/z}^2}{N-r} = \frac{3066,09}{42-6} = 85,17. \quad (11)$$

The largest value dispersion is the one generated by the group factor. According to requirements of the law of normal distribution, data can be verified by means of Fisher criterion, taking into consideration the degree of freedom in calculation of adjusted dispersions:

$$F = \frac{s_{y/x}^2}{s_{y/z}^2} = \frac{534,45}{85,17} = 6,28. \quad (12)$$

Conclusion

The F calculated coefficient is compared with the table coefficient in order to verify the degree of significance of the group factor. If $F_{calc}>F_{tab}$, then the connection is significant and the characteristic of the grouping is determinant for the variation of the result characteristic.

For a probability of 0.99, there is reference in the F distribution table and for $n_1=5$ and $n_2=36$ is the proximate value of F=3.70. Thus $F_{calc}>F_{tab,}$ 6.28>3.70, so we may say the age of the tourists arrived on the seaside for balneary tourism is a factor that influences vacation duration.

We may also overrule the null assumption and accept the alternative theory that indicates the existence of connection between the two variables (dependence).

References

Biji M., Elena Maria Biji, Eugenia Lilea, C.Anghelache. 2002. *Tratat de Statistică.* Ed. Economică, Bucharest, pp. 315-316.

Biji Elena Maria, Mirela Secară, Gratiela Popa.2006. *Statistică.* Editura Ex Ponto, Constanţa.

Biji Elena Maria, Eugenia Lilea, P. Wagner. 1996. *Statistică teoretică şi economică***.** Editura Didactică şi Pedagogică, Bucharest.