ANOVA in marketing researches

Tomescu Dumitrescu, C.

October 2005

ANOVA in marketing researches – Case study on the Chevrolet cars' market

In order to make an analysis of the variation regarding ANOVA, we considered the commercial society S.C. RUCOM S.A. from Craiova, Dolj, 107 Caracal Street, dealer in the sale of cars mark Opel, even from 2002.

Starting with 2004, S.C. RUCOM S.A. becomes dealer for cars mark Chevrolet too, cars that first appear on the market in Oltenia that year, being, as a matter of fact, not that known on the Romanian market, in general.

The Chevrolet cars, being them from the variety Kalos, Lacetti or Evanda, but especially the ones from the variety Kalos, present a few attractive characteristics for the potential clients. Thus, in the first place, the Chevrolet Kalos cars are more convenient, more advantageous, from more points of view, than many of the cars that presently exist on the market. Secondly, being produced by a firm famous for the quality of its products, respectively General Motors, the quality of the Chevrolet cars is at least as high as the one of other foreign cars that are presently sold on the market. Thirdly, the sale price of these cars is lower or, at most, equal to the one of similar cars.

The marketing director, in order to launch on the market the new product, and we refer here to Chevrolet Kalos, must take a decision in what concerns the marketing strategy. Thus, he considers that the Chevrolet Kalos car could be advertised by emphasizing both the advantage that this one presents and the quality and the price that characterizes it. In order to make the marketing strategy regarding advertising a real success, it is important to be aware of the extent to which every single one of these three more important characteristics of the product influences the purchase decision of the potential clients. In this respect, the marketing manager of S.C. RUCOM S.A. Craiova wants to make an experiment. The respective experiment presupposes launching on the market and

selling the Chevrolet Kalos car in three different counties, respectively Dolj, Gorj and Olt, through a publicity axed on a single characteristic, different from one county to another. Thus, in the county Dolj, the publicity puts the accent on the high quality by which the Chevrolet cars, in general, and the Chevrolet Kalos variety, in particular, are characterized, quality fully proved through reliability, through the special characteristics of the motorization, the reduced consumption of fuel etc. In Gorj, the marketing strategy maps out realizing publicity by putting the accent on convenience, on the advantage of the new model Chevrolet and refers here to its increased comfort, to its standard and optional endowments, to the design very pleasant for the eye etc. In Olt, the publicity made to the new product of the firm General Motors, is axed on the not that high price of sale of this one as compared to the price of other cars. The number of cars sold monthly is registered for 12 consecutive months, respectively during 2004, at the same time with the release of the campaign.

Through the agency of this campaign, the marketing manager of S.C. RUCOM S.A. wants to observe if there are differences between the three strategies of advertising in order to conceive a corresponding marketing strategy for the following period in what concerns the sale of the Chevrolet Kalos cars.

At the end of the campaign, after those 12 experimentation months passed, the results reflected in the table no. 3.7 are obtained.

Table no. 3.7. The monthly sales in those three counties

Dolj (quality)	Gorj (advantage)	Olt (price)
3	2	4
6	5	2
7	5	5
4	3	6
6	7	6
7	8	7

10	6	9
6	4	8
15	10	14
8	6	8
9	9	7
16	12	16

Taking into consideration the fact that we speak here about quantitative data, the objective problem that is brought forward is the comparison of the three populations considered, respectively the sales of Chevrolet Kalos cars in the three counties.

The statistic method that we will use in this respect is part of the descriptive statistic and it's about the analysis of the variation, more precisely ANOVA with a single factor.

In order to apply this method, the formulation of the work hypotheses is imposed, respectively the null hypothesis, that is about to be verified, and the alternative hypothesis.

H_0: *there are no differences between the car sale averages from the three counties.*

H_1: *at least two averages are different from one another.*

The response variable is concretized in the monthly sales and the experimental unity is represented by the months when we register the sales' figure in those three counties. The registered sales' figure represents the responses. As we can observe, there is a single factor, the approach of publicity, that defines the populations and there are three levels of this factor. These levels are:

- publicity by emphasizing the advantage, the convenience;
- publicity by emphasizing the quality;
- publicity by emphasizing the price.

By applying the variation analysis method, more explicitly presented in subchapter 3.2. of this chapter, we will proceed to the effectuation of the calculi in order to obtain the value of the signification test F of Snedecor.

According to the relation (3.30), $F = \dfrac{MST}{MSE}$.

We will calculate the averages of the samples and the general average and we will obtain:

$$\overline{x}_1 = 8,08; \overline{x}_2 = 6,42; \overline{x}_3 = 7,67$$

$$\overline{\overline{x}} = 7,39$$

Therefore, the value of SST, respectively the factorial variation, will be:

$$SST = \sum_{j=1}^{k} n_j \left(\overline{x}_j - \overline{\overline{x}}\right)^2 = 12(8,08-7,39)^2 + 12(6,42-7,39)^2 + 12(7,67-7,39)^2 = 18,05$$

In our example, SST=18,05. Is this value high enough in order to indicate that the averages of the populations are different from one another?

By calculating the sample dispersions for our example, we will obtain:

$$s_1^2 = 14,4; s_2^2 = 7,91; s_3^2 = 14,22$$

As a consequence, SSE, respectively the residual variation, will have the following value:

$$SSE = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + (n_3 - 1)s_3^2 = 11\cdot14,4 + 11\cdot7,91 + 11\cdot14,22 = 438,5$$

In this precise case, the number of liberty degrees is:

v_1=k-1=3-1=2 and v_2=n-k=36-3=33.

The two averages of the squares calculated for the data of the considered example, respectively the sales of Chevrolet Kalos cars by S.C. RUCOM S.A. Craiova, have the following values:

$$MST = \frac{SST}{k-1} = \frac{18,05}{2} = 9,028; MSE = \frac{SSE}{n-k} = \frac{438,5}{33} = 13,29$$

$$F = \frac{MST}{MSE} = \frac{9,028}{13,29} = 0,679$$

For a level of signification of $\alpha = 0,05$, $F_{0,05,\ 2,\ 33}$=3,285.

According to the relation (3.10), through comparison we will obtain:

$$F_{calculated} = 0,679 < F_{0,05.2.33} = 3,285$$

and we can translate this through the fact that the variation between groups (SST) is not high enough in order to be able to reject the null hypothesis. As a consequence, the null hypothesis is accepted and the alternative hypothesis is rejected, and this means, in fact, that the differences between the averages of the treatments don't differ a lot from one another.

Consequently, it's evident that the car monthly sales' averages don't differ significantly for the three counties considered.

According to the results obtained by applying ANOVA, we can state that the three considered characteristics of the Chevrolet Kalos cars altogether are equally appreciated by the buyers, this meaning that the marketing manager of S.C. RUCOM S.A. Craiova must prepare an advertising strategy that will equally emphasize both the advantage, the quality and the price of the respective cars.

The results for the considered example are to be found in table no. 3.8.

Table no. 3.8. Table ANOVA for S.C. RUCOM S.A. Craiova

The source of variation	The degrees of liberty	The sum of the squares	The average of the squares	F
Treatments	2	18,05	9,028	
Error	33	438,5	13,29	F=0,679
Total	35	456,55		

Using the program EXCEL for WINDOWS, the results from the screen above are obtained.

Excel screen showing toolbar, cell B1 = "Anova: Single Factor"

	B	C	D	E	F	G	H	I
1	Anova: Single Factor							
2								
3	SUMMARY							
4	Groups	Count	Sum	Average	Variance			
5	Avantaj	12	97	8.083333333	15.71969697			
6	Calitate	12	77	6.416666667	8.628787879			
7	Preț	12	92	7.666666667	15.51515152			
8								
9								
10	ANOVA							
11	ce of Varia	SS	df	MS	F	P-value	F crit	
12	Between G	18.05555556	2	9.027777778	0.679399468	0.513870205	3.284924333	
13	Within Gro	438.5	33	13.28787879				
14								
15	Total	456.5555556	35					
16								
17								
18								
19								

In the first table of this screen we can find data referring to the number of observations for every population, the sum of their values, as well as the averages of every single one in part.

In the second table we can find the table ANOVA with a single factor on the independent samples.

In order to process the data with this program, the following steps are to be covered:

- the data are introduced in the work register of the program EXCEL;
- the menu Tools is accessed from the menu bar;
- the option Data Analysis is chosen from the Tools' menu and, from the dialogue window, the Anova : Single Factor is chosen;
- the requested conditions must be specified in this option's dialogue window, respectively Input Range and Labels, if it's necessary.

As well, the use of the program Excel, made also possible the construction of the bar charts for every single one of the three levels of the factor "marketing approach", as it is shown in the figure (3.1), (3.2), and (3.3), by accessing the menu Insert from the menu bar and then by choosing the option Diagram.

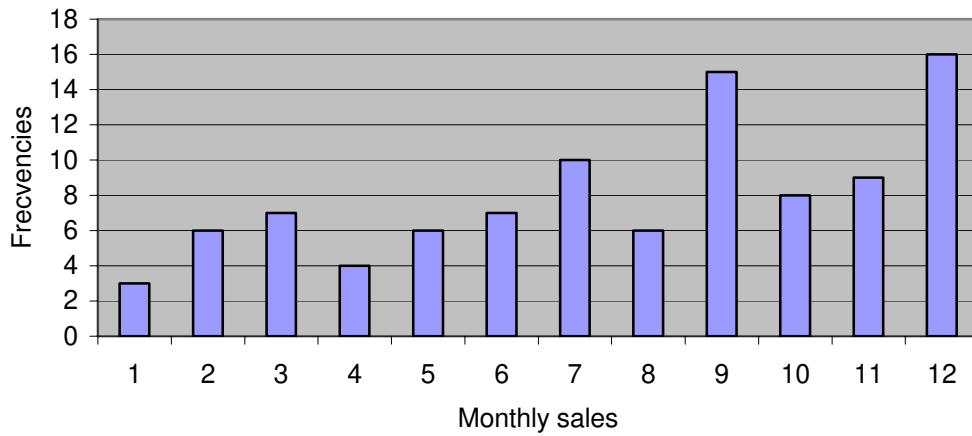Figure no. 3.1. The bar chart of the monthly sales in Dolj

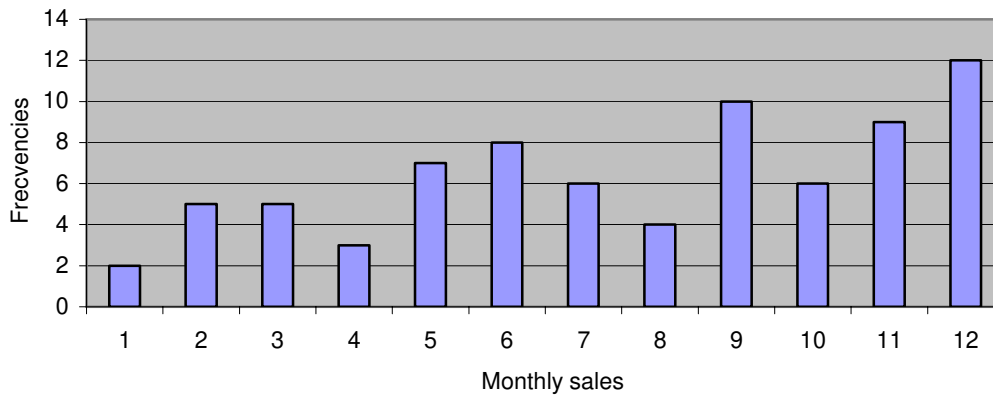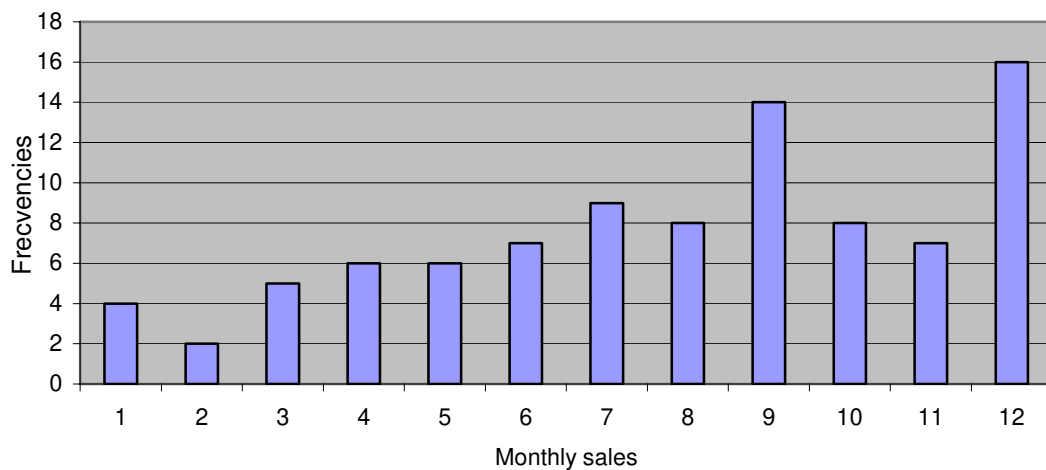Figura nr. 3.2. The bar chart of the sales in Gorj

Figura nr. 3.3. The bar chart of the sales in Olt

The marketing manager of S.C. RUCOM S.A. Craiova also wants to know if the publicity realized through the agency of television and the one

realized through the agency of the written press, besides the fact that it is concentrated on different aspects of the product, lead to different levels of cars' sale.

But the question that the marketing manager asks himself is: Which of the two media in which he intends to promote his products, respectively the television or the newspapers, is more viable?

Besides the change of the marketing approach, S.C. RUCOM S.A. decides to advertise through the agency of television or newspapers. As a consequence, the experiment is repeated, in parallel with the previous one, but in a different way. Thus, the marketing manager selects six counties: Dolj, Gorj, Mehedinti, Valcea, Bucharest. In Dolj, they decide to emphasize the convenience of the Chevrolet Kalos cars through publicity and all the publicity is made through the agency of television. In Gorj, the advertising campaign emphasizes the quality of the product and the publicity is directed towards newspapers. In Olt, the publicity emphasizes the price and is realized through the agency of television, while in Mehedinti the publicity is realized through the agency of written press and emphasizes the advantage convenience of the new product. In Valcea, the publicity is axed on the quality of the product and is totally directed towards television and in Bucharest the low price of the Chevrolet Kalos cars is emphasized by publicity and this thing is realized through the agency of the written press, more precisely of the newspapers.

The monthly sales, during the first ten months of 2004, were registered for every county in part, as shown in table no. 3.8.

Table no. 3.8. Monthly sale of Chevrolet Kalos cars

Dolj	Gorj	Olt	Mehedinti	Valcea	Bucharest
3	2	4	1	8	0
6	5	2	3	7	2
7	5	5	5	10	1
4	3	6	4	9	3
6	7	6	6	6	0

7	8	7	2	8	2
10	6	9	3	4	1
6	4	8	9	15	2
15	10	14	4	6	0
8	6	8	12	10	3

It can be observed that for this experiment we dispose of six treatments and each of them is defined by two different factors. The first factor is represented by the "marketing approach", which has three levels (convenience, quality and price). The second factor is represented by "the media" in which the publicity is realized and has two levels, respectively the television and the written press.

Taking into consideration the fact that in this particular case we only deal with three marketing approaches and only two media of publicity, we can state that it is about *the fixed effects model*[1] of the variation's analysis.

We can solve this problem in a similar way with what we previously presented, that is we will test two work hypotheses, formulated as follows:

H_0: *All the averages for those six treatments are equal to each other.*

H_1: *At least two averages are different.*

Using the program Excel for Windows, we obtain the results presented in table no. 3.9.

Table no. 3.9. The analysis of the variation with a single factor

Anova: Single Factor

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	81.73333	2	40.86667	3.54848	0.035283	3.158846
Within Groups	656.45	57	11.51667			
Total	738.1833	59				

---

[1] Keller, G., Warrack, B., "Statistics for Management and Economics", fifth edition Dnxbury, 1999, U.S.A. , p.508.

The value of the $F_{calculated}$ statistic test is 3,548. This result brings upon even more questions, namely:

- Can we conclude that the differences of the monthly sales of cars between counties are caused by the differences between the marketing approaches? Or they are caused by the realization differences of publicity, through television or written press?

- Or maybe, the combination of the marketing approach with the media in which the publicity is realized especially determines higher or lower sales of Chevrolet Kalos cars?

Table no. 3.10. The systematization of the data depending on the two factors

Factor B: Media of publicity	Factor A		
	Convenience	Quality	Price
Television	3	8	4
	6	7	2
	7	10	5
	4	9	6
	6	6	6
	7	8	7
	10	4	9
	6	15	8
	15	6	14
	8	10	8
Newspapers	1	2	0
	3	5	2
	5	5	1
	4	3	3
	6	7	0
	2	8	2
	3	6	1
	9	4	2
	4	10	0
	12	6	3

Table 3.10 describes the mode of systematization of the data for ANOVA with two factors of influence.

As a consequence, for the present experiment we will use a complete factorial experiment where the number or the treatments is *ab* with *r* replicas on the treatment and, because *a=3, b=2* and *r=10*, results that we have *ab=6*, which means that we dispose of 10 o. Going back to our case, respectively S.C. RUCOM S.A. Craiova and its sales of Chevrolet Kalos cars, we must see to what extent the monthly sales are influenced by the marketing approach (with its three levels), by the modality in which publicity in press is realized (with its two levels), respectively the interaction of those two factors.

Thus, we will test first if there are differences between the Chevrolet Kalos cars' sale averages calculated for the three levels of the marketing approach, respectively convenience, quality, and price. In order to do this, we formulate the following hypotheses:

H_0: there are no differences between the averages of the threes levels of factor A (the marketing approach).

H_1: at least two averages are different from one another.

After making the calculi, we obtain: MS(A)=9,867 and MSE=11,356. According to the relation (3.18), we will calculate the test $F = \dfrac{MS(A)}{MSE} = 0,869$.

We can state that, for a level of signification of 0,05 with a possibility of 95% and 2 degrees of liberty, between the averages of the sales of S.C. RUCOM S.A. Craiova for the three levels of the marketing approach there are no significant differences because $F_{calculated}=0,869 < F_{tabelar}=3,168$.

In order to see if factor B, that is the type of press through the agency of which publicity is realized, affects the sales of cars, we formulate the following work hypotheses:

H_0: there are no differences between the averages of the factor's two levels.

H_1: those two averages are different from one another.

As a consequence of making the calculi, we obtain: MS(B)=91,267 and MSE=11,356. By applying the relation (3.31), we obtain a value of F=8,037.

Because, for a level of significance of 0,05 $F_{calculated}=8,037>F_{tabelar}=4,019$, we reject the null hypothesis and we accept the alternative hypothesis, according to which, between the averages of the two levels of factor B there are significant differences. In other words, the level of the sales of Chevrolet Kalos cars at S.C. RUCOM S.A. Craiova is different when publicity is realized through the agency of television, as compared to the case in which it's realized through the agency of newspapers.

In order to verify if the interaction of the two factors influences the sales of the considered society, we formulate the work hypotheses as follows:

H_0: *the two factors don't interact in influencing the average of the monthly sales.*

H_1: *the two factors interact in influencing the monthly sales' average.*

In this situation we obtain the following values: MS(AB)=29,867 and MSE=11,356. As a consequence, F=2,63.

Because in the conditions of probability mentioned, $F_{calculated}=2,63<F_{tabelar}=3,168$, we accept the null hypothesis and reject the alternative hypothesis, therefore the factors A and B don't interact in influencing the monthly sales' average. Therefore, the Chevrolet Kalos cars' sales are not modified under the concomitant action of the marketing approach and of the modality of realizing publicity.

Processing the data with the program Excel for Windows supposes going through the same steps as in the case of the analysis of the variation with one factor, with the difference that from the dialogue window Data Analysis the option Anova: Two Factor With Replication is chosen.

	A	B	C	D	E	F	G
4							
5	Count	10	10	10	30		
6	Sum	78	56	72	206		
7	Average	7.8	5.6	7.2	6.866666667		
8	Variance	9.288888889	2.488888889	9.511111111	7.498850575		
9							
10							
11	Count	10	10	10	30		
12	Sum	48	58	26	132		
13	Average	4.8	5.8	2.6	4.4		
14	Variance	17.51111111	6.177777778	23.15555556	16.3862069		
15							
16	Total						
17	Count	20	20	20			
18	Sum	126	114	98			
19	Average	6.3	5.7	4.9			
20	Variance	15.06315789	4.115789474	21.04210526			
21							
22							
23	ANOVA						
24	Source of Variation	SS	df	MS	F	P-value	F crit
25	Sample	91.26666667	1	91.26666667	8.037181996	0.006435086	4.019540256
26	Columns	19.73333333	2	9.866666667	0.86888454	0.425199556	3.16824611
27	Interaction	59.73333333	2	29.86666667	2.630136986	0.081285453	3.16824611
28	Within	613.2	54	11.35555556			
29							
30	Total	783.9333333	59				
31							

In the first table of this Excel screen we find systematized data referring to the number of the replicas for each of the three levels of factor A (the marketing approach), the size of the samples for each level, but also their averages and all these aspects and for the entire factor in general, however only corresponding to the first level of factor B. In the second table we can find the same type of information, just that in correspondence with level two of factor B. In the third table we can find systematized the data of the first two tables, respectively the number of the replicas, the size of the samples and the averages for the levels of factor B.

In the first table of the Excel screen we can find the data systematized in an ANOVA table with two factors and independent samples.

The process of the data can be realized with other packages of software programs also, and one of them broadly used in the statistic analysis of the data, is the package SPSS for Windows as well, Version 10.0. The results obtained with this one are systematized in a similar way with the ones from Excel.