TAKING DECISION BASED ON THE REGRESSION MODEL USING EXCEL 2016

Ana-Maria Mihaela IORDACHE [1*]

ABSTRACT

The management decision is represented by the process of choosing a line of action in order to achieve goals, the application of which affects the activity of at least one other person than the decision-maker. The decision making and the possibilities of assisting it depend on the context in which this process takes place and on the typology of decisions. This paper discusses how to make a decision about the choice of the linear regression model. The case study is focused on the observations made on a production company for a period of 30 years. At the end of the study the regression model was chosen in such a way that maximizes the possibility of making the right decision on the production plan over the following periods.

KEYWORDS: *decision, Excel 2016, regression, information, linear, knowledge*

INTRODUCTION

The decision is the result of conscious activities and means choosing an optimal solution to solve a problem, which has implications in the activity of another person. The decision involves choosing a course of action and engaging in it, which usually involves the allocation of resources. The decision belongs to a person or group of persons who: have professional competence, have the necessary authority and are responsible for the use of resources in certain given situations.

The decision is the focus of management activity because it is found in all its functions. The success of a decision depends on few factors as: the decision quality, the way of implementing the decision and how the objectives proposed by its implementation have been achieved. Moreover, the business integration within the business environment depends on the quality of the decision. The quality of the decision refers to its compatibility with the existing restrictions.

The essential attribute which characterizes the decision is that of choosing from more alternatives. The alternatives should only be identified from an existing offer (for example, the list of job placement companies or products on the market) or invented by making scenarios (for example, establishing the production program variants).

When companies make decisions, they must take into account the following factors that influence their quality:

[1*] Lecturer, Phd, Romanian-American University, Bucharest, iordache_ana_mari_mihaela@yahoo.com

Decision=f[(fc,fin);(V,M,R)], V=f(C,Q), where

- ***fc*** are the known factors (information, restrictions, influences);
- ***fin*** are the factors of uncertainty and risk determined by the environment;
- ***V*** is the value of the human factor;
- ***C*** represents the person's knowledge and experience who makes the decision;
- ***Q*** represents the ability of the person to adapt who makes the decision;
- ***M*** is the motivation of the decision;
- ***R*** is the responsibility of the person who makes the decision.

Through criteria that every manager should take account when make a decision, are: avoiding improvisations and subjectivism in the process of taking decisions, the decision must be taken only by persons who are legally invested and authorized to do so, the decision must include all the elements necessary for proper understanding and implementation, it is preferably a good decision that is taken in time than a very good decision taken later, it is intended to get the best effect for a certain effort, to ensure that decisions on the various departments of the organization are compatible with each other and lead to accomplishment the overall objective of the enterprise.

Taking a decision can be treated from two perspectives: classic and knowledge. From the classic perspective, a decision is treated as a choice of a possible alternative to the action from several possible pre-identified ones. In general, the number of alternatives identified and taken into account in the decision-making process is high. The issues that arise here relate to the number of alternatives, how to identify the hidden alternatives, and how these alternatives are managed to be forgotten or omitted. Then there is the problem of selecting one of them, following an analysis of the alternatives, highlighting the implications of each and the impact in accordance with the proposed objectives. The vision from the perspective of knowledge treats the decision as knowledge of how the action is carried out. When a decision is treated as knowledge, the decision-making process becomes a process of creating new knowledge through the use of existing ones. Decision making is seen from this perspective as an intensive activity in knowledge.

Microsoft Excel is an outstanding software platform of MS Office. It has great applications in statistical analysis in every field wherever it is needed statistical evaluation for assessing significance of research pursuits or control of quality. Microsoft Excel application has inbuilt provision for designing all sorts of graphics for presentation of data. The data can be arranged in tabulated forms in ascending or descending order, we can find out the average (mean) and standard deviation, and do statistical analysis and verify the results of statistical analysis done by manual methods of statistical analysis. Calculations can be done quite quickly and accurately using Excel worksheet. There are built-in functions to calculate percentage points or p-values for many distributions of interest in statistical models.

THE REGRESSION MODEL

Regression is a method of statistical analysis by which determined values are assigned for a dependent variable (y) and for one or more independent variables xj (j = 1, ..., m), we seek a simple expression of the function that expresses the connection between them. By extrapolation, based on this expression and other values of the independent variables, we can look for corresponding values of the dependent variable. If the statistical data expresses an evolution over time, then the calculation of the future values of the dependent variable y is called prognosis and the one of the past values - retrogression.

In the regression analysis, the dependent variable y is considered to be a function of one or more independent variables, as shown below:

$$y = f(x_1,x_2,.....x_n)+\varepsilon,$$

where ε represents the deviations of the theoretical variable from the experimental variable.

In the regression analysis, it is desirable to determine the function f, so that ε is a variable with the mean of zero and the minimum dispersion.

If this pattern is not satisfactory, you can use the multi-linear regression that has the form:

$$y = a_1x+a_2x^2+a_3x^3+...a_mx^m+b$$

In multi-line regression, the coefficients a_j $(j=1,....,m)$ and b must be determined.

When m = 1 the dependence is linear and the regression function is represented by a straight line, whose slope is given by a1, and b gives the ordinate of the point of intersection of the straight line with the O_y axis.

There are various other dependency functions that can be reduced to multi-linear function by some transformations. For regression, the least squares method is most often used to minimize the squares of the distances between the values given for y and those calculated by the regression function.

CHOOSING THE RIGHT REGRESSION EQUATION

Various regression functions, such as polynomial functions, can be reduced by appropriate transformations to the multi-line regression function. Regression can also be done with Regression in Data Analysis. But the question is whether a regression function is good or not, how good it is and to what extent is it better than another. These assessments can be made using statistical indicators that cover either the whole of the function or its components. We used in the calculations and interpretation of statistical indicators R^2, F and t (Student).

The correlation coefficient R is between -1 and 1. If R = 0, there is no correlation between the respective variables in the regression model, and if R = ± 1, the correlation is perfect. If all the values are on the regression curve, it results that R^2=1 and therefore data matching is perfect for the calculated regression curve. So, the closer R^2 is to 1, the more regression equation is more appropriate.

Test F indicates whether the entire regression equation is significant or not. In Excel we get a calculated value of F, which we note with F_c. The theoretical value of the inverse statistical distribution F is noted by F_t and it has values determined in the statistical tables in the specialized books. If $F_c > F_t$, , means that the regression equation is significant.

Test *t* (Student) for all coeficients a_j $(j=1,2,....m)$ of the variables and for the free term b in the regression equation indicates whether or not the relevant variables or the free term are significant.

It starts with linear regression. If at least one of the three indicators does not indicate a good correlation, a quadratic regression (regression grade equation 2). If the square regression is good, we will also try a cubic regression and determine which is the most significant.

The effective comparison of regressions will be done in 4 stages: data preparation for polynomial regressions; linear regression; quadratic regression; cubic regression.

1. Preparing data for polynomial regression

Consider the case of evolution of an organization on a number of 30 years. The annual production value in € billions represents the independent variable x, and the production costs in € billions represent the y dependent variable in the model. In the worksheet we will first type the years and production x. In the adjacent columns we calculate x_2 and x_3 and then type the y costs in the next column (Table 1). This column fill order is mandatory, because in Regression the columns with data for independent variables must be contiguous.

Table 1. Statistical data of the company

	A	B	C	D	E
1	The company costs evolution				
2	Observation number	Production volume x1	x2 (=x1^2)	x3=(x1^3)	Cost productio
3	1	5.386	29.009	156.242	1094.15
4	2	8.128	66.064	536.971	3115.36
5	3	6.161	37.958	233.859	2307.32
6	4	6.217	38.651	240.294	2145.15
7	5	10.265	105.370	1081.625	5157.27
8	6	7.870	61.937	487.443	1751.86
9	7	8.212	67.437	553.792	2412.08
10	8	7.382	54.494	402.274	2350.33
11	9	7.935	62.964	499.621	2664.81
12	10	4.925	24.256	119.459	3054.94
13	11	1.254	1.573	1.972	1078.63
14	12	2.010	4.040	8.121	3071.48
15	13	6.235	38.875	242.387	2199.07
16	14	4.254	18.097	76.983	1984.25
17	15	8.524	72.659	619.342	2209.25
18	16	9.214	84.898	782.248	3032.66
19	17	11.235	126.225	1418.140	4634.34
20	18	4.525	20.476	92.652	2101.53
21	19	6.444	41.525	267.588	2295.63
22	20	10.235	104.755	1072.170	3201.67
23	21	10.592	112.190	1188.321	2789.65
24	22	12.654	160.124	2026.206	3488.61
25	23	4.235	17.935	75.956	2000.95
26	24	4.658	21.697	101.064	2648.15
27	25	9.884	97.693	965.602	2732.38
28	26	16.235	263.575	4279.144	5841.19
29	27	4.685	21.949	102.832	2000.65
30	28	3.958	15.666	62.005	2011.01
31	29	7.199	51.826	373.093	2125.95
32	30	14.856	220.701	3278.730	7317.62

2. Linear regression

Running regression for the equation *y=ax+b*.

Choose Data menu, then Data Analisys and finally press Regression. The Regression window will appear and then the user must establish some parametres for the regression model, like: the input y range, input x range, where the output results should be available (in a range, in another worksheet or another workbook).

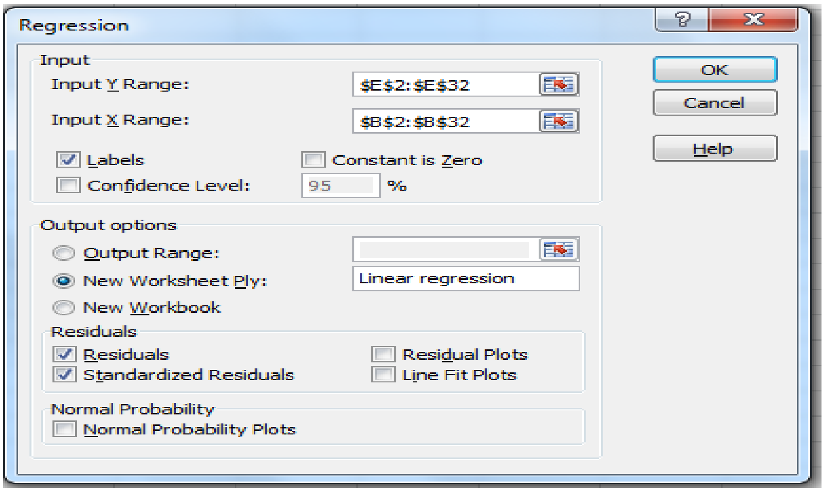

Figure 1.

From the *Summary Output* report generated in the worksheet it results that:

- a=298.48, b=584.98
- n=30 from *Observations*
- *k=2 in the first line of the column df.*

It is noted that R^2=0.602 (*R-Square line*), F_c=42.42 (*F*), t_a=6.299, t_b=1.546 (from *t-Stat*). From $df_1=k-1=1$ and $df=n-k=df_2=28$, it results F_T=FINV(0.05,1,30)=4.17 and t_T=TINV(0.05,30)=2.04. The *F* and *t* tests are good, but R^2 does not indicate a good correlation. Next we will try a grade 2 regression equation (square).

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.77616633							
R Square	0.60243417							
Adjusted R Square	0.58823539							
Standard Error	859.254703							
Observations	30							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	31325817.89	31325818	42.42859	4.64673E-07			
Residual	28	20672922.06	738318.6					
Total	29	51998739.95						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	584.982566	378.3009315	1.546342	0.133251	-189.9317637	1359.897	-189.93176	1359.8969
Production volume x1	298.484086	45.82388325	6.513723	4.65E-07	204.6181163	392.3501	204.618116	392.350056

Summary Output and ANOVA for linear regression

3. Squared regression

The regression equation is $y = a_1x+a_2x^2$.

The results are obtained in the sheet to which we assign the name SQUARED.

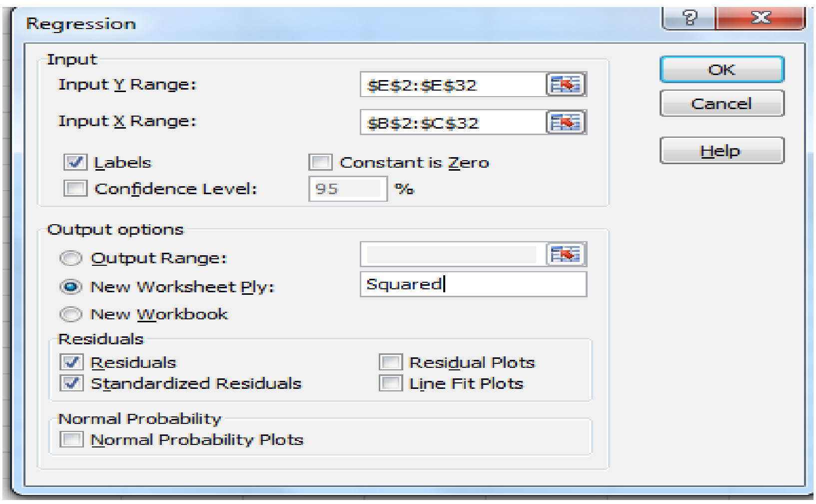

Figure 2.

In the report from the square sheet is obtained: a_1=181.40, a_2=28.18, b=2269.039, R^2=0.846, F_c=34.017, t_1=-1.198, t_2=3.28, t_b=3.73.

From $df_1=k-1 = 2$ and $df=n-k=df_2=27$ it results that F_T=3.35 and t_T=2.05. It is found that R^2 is much better than linear regression, and F and t tests are good.

It results that the second order regression is good, but we will try cubic regression too.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.846104895							
R Square	0.715893494							
Adjusted R Square	0.694848568							
Standard Error	739.6991462							
Observations	30							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	2	37225559.62	18612780	34.01739	4.1879E-08			
Residual	27	14773180.33	547154.8					
Total	29	51998739.95						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*Upper 95.0%*
Intercept	2269.039374	607.5184212	3.734931	0.000888	1022.514538	3515.564	1022.515	3515.56421
Production volume x1	-181.4009325	151.372734	-1.19837	0.241185	-491.9921275	129.1903	-491.992	129.1902625
x2 (=x1^2)	28.18547389	8.583491045	3.283684	0.002835	10.57360502	45.79734	10.57361	45.79734275

Summary Output and ANOVA for the second order regression

4. Cubic regression

The cubic regression equation has the following form:

$$y = a_1x+a_2x^2+a_3x^3+b$$

The regression results are obtained in the sheet we named Cubic.

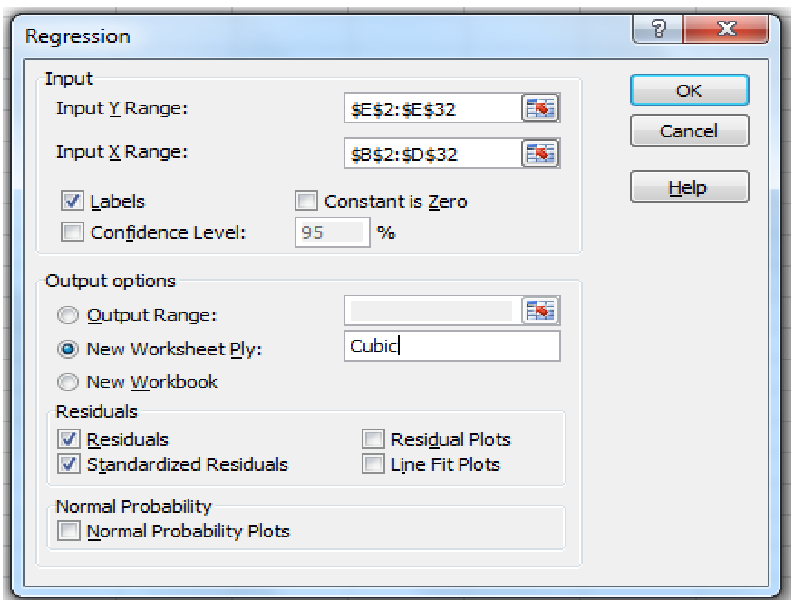

Figure 3.

From the obtained report we get the following values: a_1=239.58, a_2=36.07, a_3=0.303, b=2384.014, R^2=0.84, F_c=21.86, t_1=0.55, t_2=0.65, t_3=0.145, t_b=2.37. Because df_1=3, and $df=df_2$=26, it results that F_T =2.97, t_T=2.055. R^2 is much better than linear regression and second order, the F test is good, but the t test is not good for $a3$. Finally, we can drop the cubic equation.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.846241145							
R Square	0.716124076							
Adjusted R Square	0.683369161							
Standard Error	753.4839686							
Observations	30							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	3	37237549.58	12412517	21.8631	2.79149E-07			
Residual	26	14761190.37	567738.1					
Total	29	51998739.95						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2384.014926	1004.447734	2.373458	0.025297	319.3430393	4448.6868	319.343039	4448.68681
Production volume x1	-239.586478	429.0520304	-0.55841	0.581343	-1121.51556	642.3426	-1121.5156	642.342601
x2 (=x1^2)	36.07760159	55.00676994	0.655876	0.517667	-76.9904334	149.14564	-76.990433	149.145637
x3=(x1^3)	-0.303219054	2.086515549	-0.14532	0.885576	-4.59211319	3.9856751	-4.5921132	3.98567508

Summary Output and ANOVA for cubic regression

After performing the regression analysis, the following operations can be done in Excel: calculation of retrognosis and prognosis, as well as graphical representation of the initial data and the obtained results

These operations can be done in three steps: writing, generating and arranging data in a proper order; graph generation using ChartWizard; interactive adaptation of chart elements with appropriate dialog boxes.

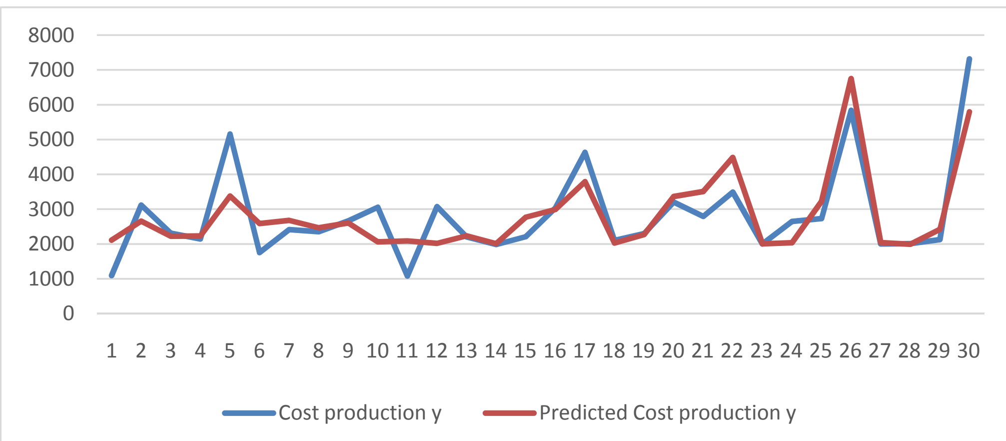

Figure 4.

The graphical represented of cost of production and predicted cost of production

CONCLUSIONS

The successful implementation of a decision is about avoiding conflicts of interest, understanding the decision by everyone who has to execute it. The criteria that every manager should take account when make a decision are: the scientific substantiation, the legality, the completeness, the opportunity and the efficiency of a decision. The context in which the decision is made may influence the nature of the decision-making, so implicitly how a decision-support system can act. It can be analyzed from several points of view, like: the decisional level, the degree of maturity, the degree of competition and the structure of the organization. The decision-making process is based on information. The information is necessary to: define and structure the problem, explore and choose between alternative solutions, review the effects of the implemented choice, solve the problems. To make good decisions, enterprises have a permanent desire to understand the reasons for certain values of business measures.

REFERENCES

[1] Virgil Chichernea, Sisteme suport de decizie, Ed. PrintGroup, 2010;

[2] Vasilica Voicu, Sisteme suport pentru adoptarea deciziei, Editura Universitară, Bucureşti, 2010;

[3] Wayne Winston, Microsoft Excel Data Analysis and Business Modeling (5th edition), Microsoft Press, 2016, ISBN: 978-1-509-30422-6

[4] Thomas J. Quirk, Excel 2016 for Engineering Statistics, Springer, Cham, Switzerland, 2016, ISBN: 978-3-319-39181-6

[5] Michael Alexander, John Walkenbach, Richard Kusleika, Excel 2016 Formulas Mr. Spreadsheet's Bookshelf, John Wiley & Sons, 2016, ISBN: 978-1-119-06786-3

[6] Jeff Sauro, James R Lewis, Quantifying the User Experience: Practical Statistics for User Research, Morgan Kaufmann, 2016, ISBN: 978-0-128-02548-2

[7] Rayat C.S., Applications of Microsoft Excel in Statistical Methods. In: Statistical Methods in Medical Research. Springer, Singapore, 2018, ISBN: 978-981-13-0827-7

[8] Andre Petermann, Graph Pattern Mining for Business Decision Support, Database Research Group University of Leipzig, https://dbs.uni-leipzig.de/file/paper03.pdf

[9] Steven Weikler, Office 2016 for Beginners – The perfect guide on Microsoft Office, 2016, ISBN: 978-1537205755

[10] Bill Jelen, Michael Alexander, Excel 2016 Pivot Table Data Crunching, ed. Que, 2015, ISBN: 978-0-7897-5629-9