

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

Guidelines for laboratory works
on the academic discipline
"DECISION SUPPORT SYSTEMS"
for full-time students of training direction
6.030601 "Management"

Харків. ХНЕУ ім. С. Кузнеця, 2015

Затверджено на засіданні кафедри комп'ютерних систем і технологій.
Протокол № 9 від 31.03.2015 р.

Compiled by O. Zavgorodnia

G 94 Guidelines for laboratory works on the academic discipline "Decision Support Systems" for full-time students of training direction 6.030601 "Management" / compiled by O. Zavgorodnia. – Kh. : S. Kuznets KhNUE, 2015. – 60 p. (English)

Laboratory tasks and guidelines to them aiming to help acquire decision-making skills are given. The main theoretical background of the decision theory and decision support systems is provided to carry out the laboratory tasks.

Recommended for full-time students of training direction 6.030601 "Management".

Подано лабораторні завдання та методичні рекомендації до їх виконання для набуття навичок прийняття рішень. Для виконання лабораторних завдань наведено теоретичні основи теорії прийняття рішень та функціонування систем підтримки прийняття рішень.

Рекомендовано для студентів напряму підготовки 6.030601 "Менеджмент" денної форми навчання.

Introduction

Profound changes in the Ukrainian as well as global economy are evidently the sources of new business needs in quick sound decisions, which should be based on grounded analysis and forecast. Quick, real time decision-making becomes a crucial competence of every manager. Sophisticated managers do not only rely on their experience, intuition and judgment. All these elements are highly important in making decisions concerning complex systems, but they are not enough for predicting reaction to the external influences of any complex system with many variables. In such a situation prediction of the total outcome often strains human cognitive capabilities and may be daunting.

The academic discipline "Decision Support Systems" is a professionally-oriented optional discipline for students of training direction 6.030601 "Management". It is a methodic and methodological base for the choice of models and methods of effective and efficient decision-making with the help of decision support systems of various kinds. These guidelines will be helpful in the organization and support of various decisions at enterprises of any size. They are also useful for students of specialities "Administrative Management" and "Management of Innovation" in their educational activity.

The subject of the academic discipline "Decision Support Systems" is theoretical basis and methods for solving complex managerial problems with the help of decision support systems.

The aim of the academic discipline is to provide theoretical knowledge and form practical skills in using methods and techniques of decision-making. It forms the following competences: knowledge of approaches in decision theory, methods and techniques of decision-making processes mathematical support, skills in making a grounded choice of methods and decision support systems according to specific demands of business tasks and situations, etc.

Module 1

Main decision support theory terms. The notion of decision support systems

Practice 1

The decision process stages. The intelligence stage

Key goals:

to learn the decision process stages, to trace them in a real-world decision-making situation;

to gain the skills in specific business problems identification and analysis;

to learn how to build and analyze the "problem map";

to gain primary skills in alternatives list formation and evaluation.

The theoretical background (in short)

The main term we may face in decision theory is a **decision**. In general, a decision is understood as a choice from multiple alternatives (options). This choice is made by a decision-maker due to professional or vital importance. In this context a decision is seen as an act which depicts the will of an individual, his/her possibility to make any choice and influence the current situation. Making no decision is also a kind of choice, because it is a possible option which is exercised by an individual. In this case the decision is a characteristic of individual's will which lacks the result. In general, any decision is associated with a result or an outcome. So, in human perception any decision is naturally combined with choice as a will and choice as a result of some thinking procedure. In Slavic mentality decisions are also associated with a choice as a process of decision-making, it has definite linguistic ground.

The necessity of decision-making is defined by inconvenience of the present situation which is called a **problem**. A problem is a kind of situation that forces an individual to change it for the better because of inconvenience or threat under current conditions. The will to change it raises an idea or an image why or how exactly it should be changed, i.e. some **objective**. Any objective might be characterized as a desirable situation which is wished to be achieved by the decision-maker (an individual or a group). By nature, the objective might be of different levels and can be named an aim, a perspective

or a goal. **The bottom line:** *the main driver of the decision-making process is the gap between the current situation (the problem) and the desirable situation (the objective) which stimulates the decision-maker to take certain actions.*

A set of actions aimed at solving the existing problem and helping to achieve the planned objective are called **alternatives**. Alternatives mean options which exclude each other. The alternatives may represent different possible actions aimed at achieving the given objective, different hypotheses about the character of the future, different classifications and so on. The number of alternatives may range from two to an infinite number. In order to make a good choice of them, a set of criteria are used.

A **criterion** is the basis of choice, a kind of preference. Criteria are used to measure the role or extent of attractiveness of each alternative from the point of view of the objective. By nature, criteria might be qualitative and quantitative. In real business situations both types of criteria are used in order to choose the best alternative. Generally, criteria influence the choice of alternatives by way of exclusion or evaluation. Some alternatives are excluded from the set of possible options if our criterion has a characteristic of constraint (or a so-called limiting factor) [12]. In this case the criterion serves to limit the alternatives under consideration or to express the characteristic that the decision set must possess. Criteria which might be considered as factors evaluate alternatives by the range of their attractiveness in achieving the objective, they show the suitability of a specific alternative.

The procedure by which criteria are selected and combined to arrive at a particular evaluation, and by which evaluations are compared and acted upon, is known as a **decision rule**. Decision rules typically contain procedures for combining criteria into a single composite index and a statement of how alternatives are to be compared using this index.

Criteria are the measurements of effectiveness of various alternatives. Criteria correspond to different kinds of system performance. Besides economic criteria, which tend to prevail in the decision-making process within companies, it is however possible to identify other factors influencing a rational choice. They are: economic, technical, legal, ethical, procedural, political. Economic factors are the most influential in decision-making processes, and are often aimed at the minimization of costs or the maximization of profits. Options that are not technically feasible must be discarded. Legal rationality implies that before making any choice the

decision-makers should verify whether it is compatible with the legislation and with ethical principles and social rules of the company. A decision may be considered ideal from an economic, legal and social standpoint, but it may be unworkable due to cultural limitations of the organization in terms of prevailing procedures and common practice. The decision-maker must also assess the political consequences of a specific decision for individuals, departments and organizations.

Identifying decision outcomes has been a particularly difficult challenge for decision-makers as well as for decision researchers. Decision outcomes are frequently multifaceted and often difficult to fully grasp and quantitatively measure.

H. A. Simon is considered the pioneer in the development of human decision-making models, who established the foundation for human decision-making models [34]. His basic model depicts human decision-making as a three-stage process. These stages are: intelligence, design, choice. Intelligence is the identification of a problem (or opportunity) that requires decision and collection of information relevant to the decision. Design is creating, developing, and analyzing alternative courses of action. Choice is selecting a course of action from those available.

During the intelligence phase, the decision-maker observes the reality, gains fundamental understanding of the existing problems or new opportunities, and acquires general quantitative and qualitative information needed to address the problems or opportunities. In the design phase, the decision-maker develops a specific and precise model that can be used to systematically examine the discovered problem or opportunity. This model will consist of decision alternatives, uncontrollable events, criteria, and the symbolic or numerical relationships between these variables. Usage of the explicit models to logically evaluate the specified alternatives and to generate recommended actions constitutes the ensuing choice phase. During the subsequent implementation phase, the decision-maker ponders the analyses and recommendations, weighs the consequences, gains sufficient confidence in the decision, develops an implementation plan, secures needed financial, human and material resources, and puts the plan into action.

After the final choice is implemented, the decision-maker should observe the new reality and, where appropriate, follow through with intelligence, design, choice, and implementation. Moreover, phase analyses may suggest the need for revisions at the preceding phases. For example,

analyses at the choice phase may necessitate adjustments in the previous design. Such continuous monitoring and adjustment is similar to Simon's review phase.

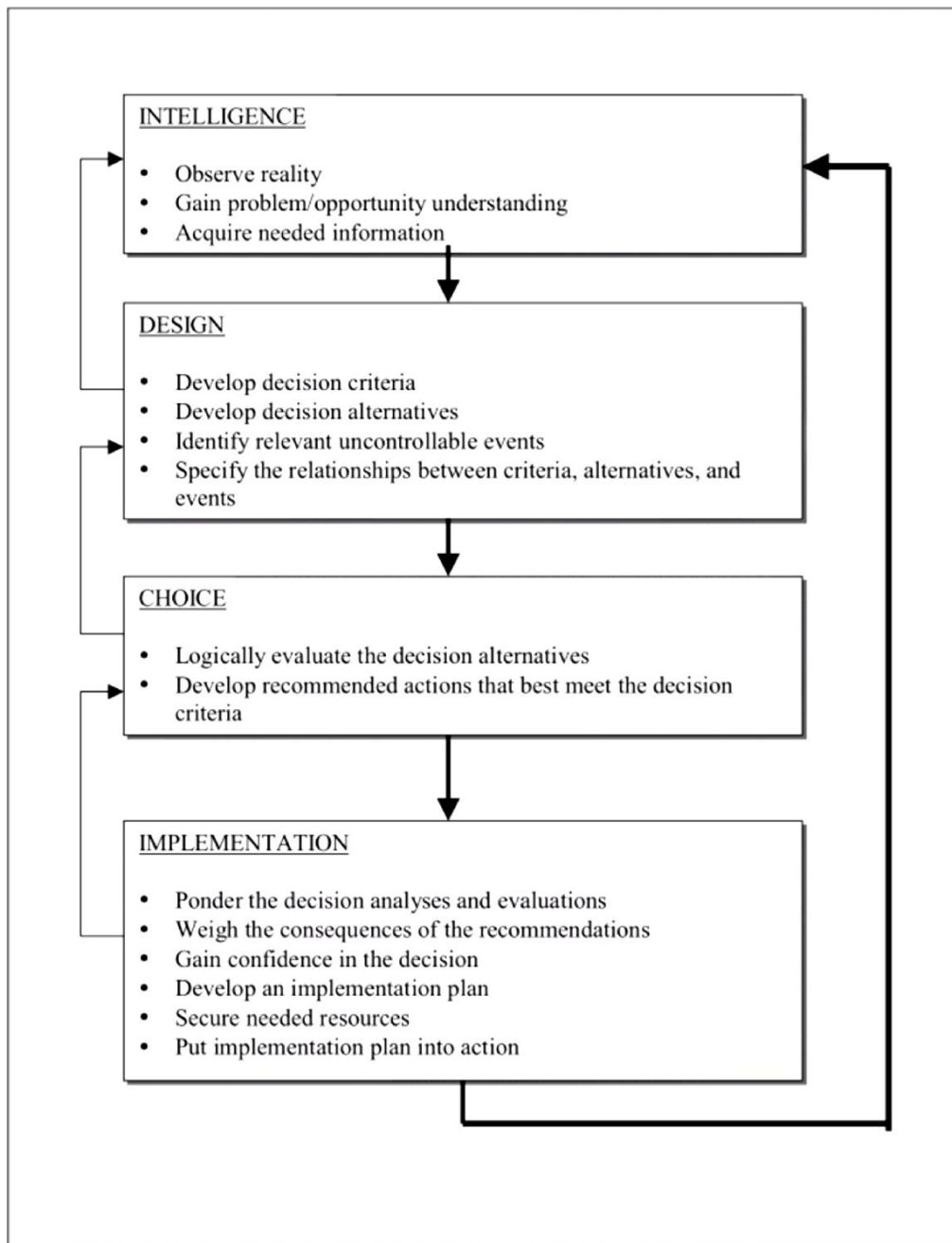


Figure 1. **The decision-making process** [34]

No decision process is this clear-cut in an ill-structured situation. Typically, the phases overlap and blend together, and there will be a return to earlier stages, as more is learned about the problem, or solutions do not work out, and so forth. Conceptually, the decision-making process applies in the same manner to individual or group decision-making.

Step-by-step guide

Step 1. In order to study the decision-making process in a real-world situation every student should choose a case from the proposed range.

Step 2. Read the chosen case and try to find out what problem the enterprise faced with (use all your knowledge of economics, operations research and other areas in order to get all the puzzles of the case situation together).

Step 3. Map the problem with the help of the decision support system Decision Explorer (made by Banxia Software Ltd).

Run the software Decision Explorer (the demo version). The first dialogue window will propose to choose the directory, where your model may be saved. You should choose a folder on disk D or S (but it would be definitely better to save your model in a folder on your flash drive), rename the model as "model_1_your second name".

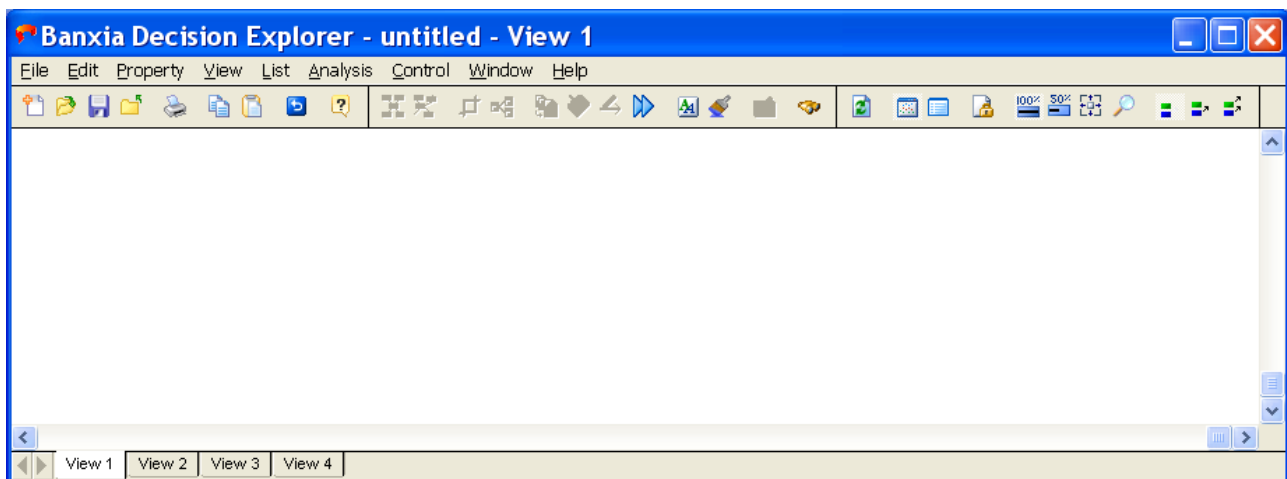


Figure 2. The start dialogue window of Decision Explorer

The Decision Explorer interface is rather easy and intuitive:

a) a double click (the left mouse button) on free (white) workspace allows making a concept-object (rectangle), where the problem or its causes may be typed in;

b) a single click (the right mouse button) on the concept (rectangle) opens a context menu. The most useful context menu options are "to move the concept", "edit it";

c) in order to change the concept look, the button "Aa" ("Define the concept style properties") should be used. It allows changing the font (its size, color, culture), the border (its shape and color);

d) in order to create connections between the concepts (arrows), the left mouse button should be pressed (the sign of an arm or question will appear). Lead it to the needed concept (the connection will be formed if the aimed concept is shaped by the dotted line).

As a result of mapping the business case problem, there might be made a model as shown in Fig. 3.

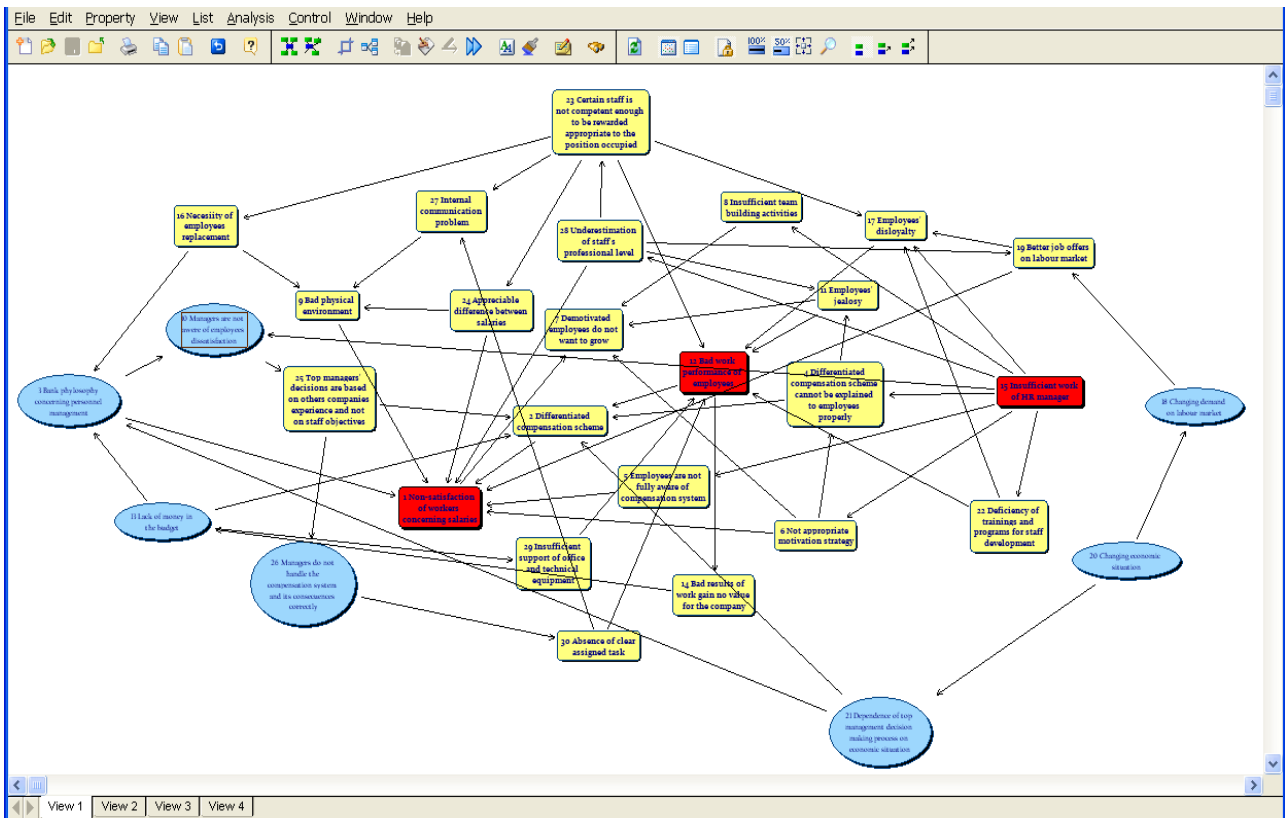


Figure 3. An example of business case problem mapping

Step 4. Save your model for the subsequent analysis.

Step 5. In order to perform analysis of the model, one should learn the buttons of the toolbar and the main menu. Special attention should be paid to the following commands:

a) View/Map Display and View/Text display – it gives an opportunity to switch between the two main modes;

b) Anaysis/Central – the most "powerful", influential concepts are ranged to the top of the list;

- c) Analysis/Domain – all the concepts are ranged according to the number of the inflowing and outflowing concepts (it is extremely useful for social tasks);
- d) Analysis/Loop – all economic multiplier loops are found;
- e) Analysis/Cluser – all clusters are depicted if they exist in this model.

Step 6. All the analysis functions should be applied to the model and all economic explanations to the results should be given.

The result of Central analysis is shown in Fig. 4.

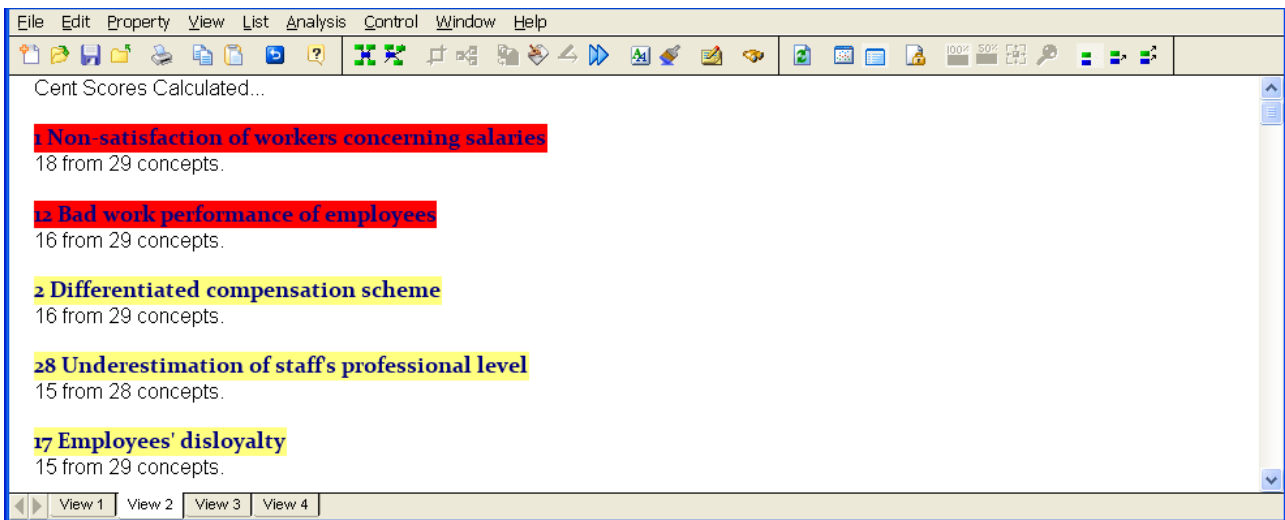


Figure 4. An example of Central analysis

The result of Domain analysis is shown in Fig. 5.

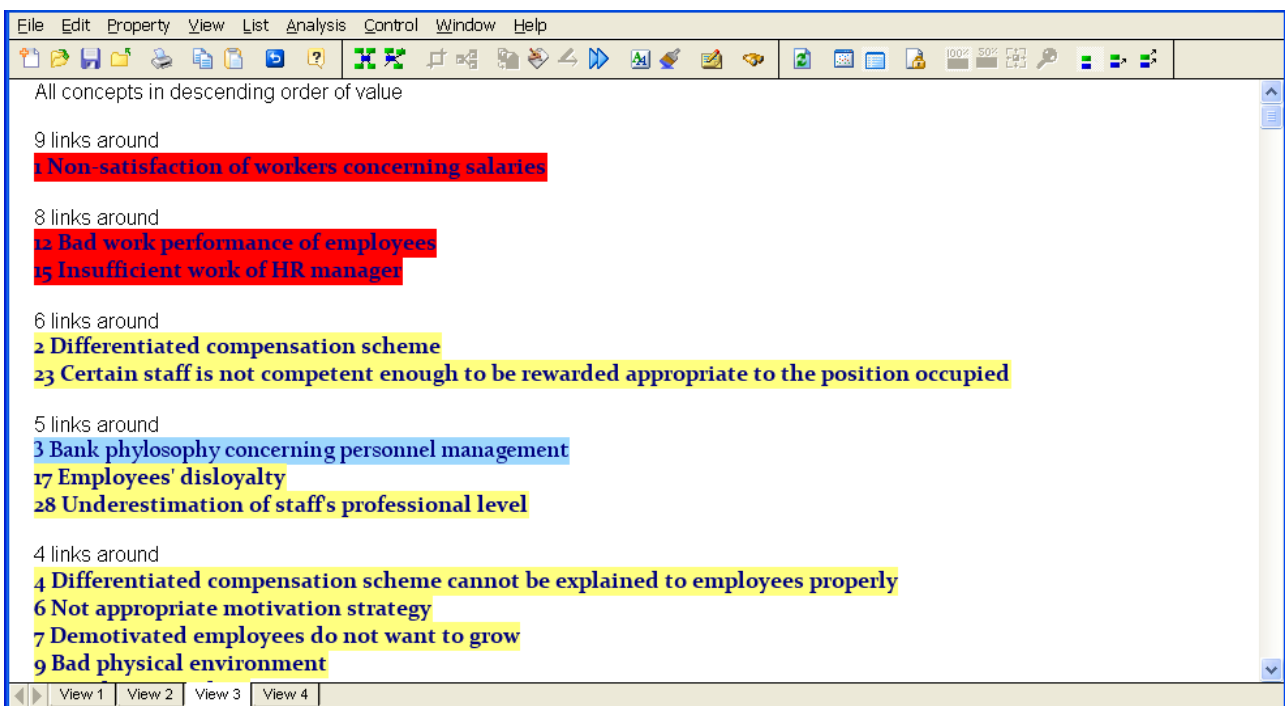


Figure 5. An example of Domain analysis

The result of Loop analysis is shown in Fig. 6.

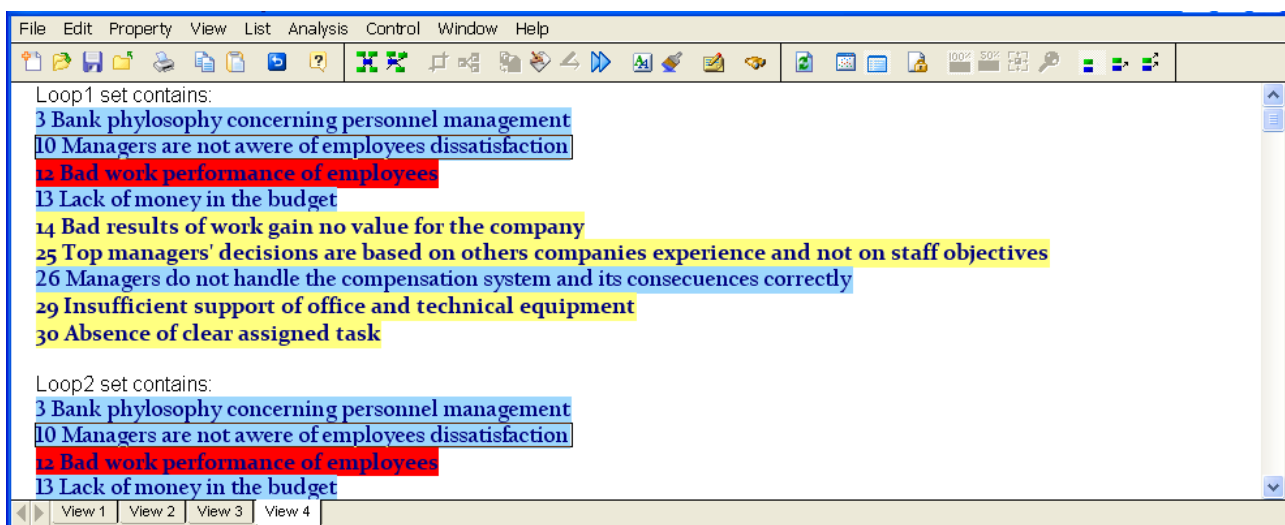


Figure 6. An example of Loop analysis

Step 7. Make a final report with all the summaries and economic explanation of the analysis results. Submit your practice report, as well as the source Decision Explorer file.

Practice 2

The decision-making process stages. Computer support of the design and choice stages of decision-making

Key goals:

to learn the decision-making process stages, to trace them in a real-world decision-making situation;

to gain the skills in solving specific business problems (the skill in comparing alternatives in multi-criteria business tasks);

to learn how to form a decision task and analyze the hierarchy of decision elements;

to gain skills in the formation and evaluation of a list of alternatives.

The theoretical background (in short)

The decision theory is a scientific discipline, which deals with two main tasks:

1) to study/explore how people make decisions;

2) to develop methods for decision-making, which help to justify the selection of the best alternative from a feasible alternative set.

As it is clear from the definition, the decision theory deals with two classes of problems. According to them, descriptive and normative theory might be distinguished.

The normative theory is a mathematical theory of decision-making, which is based on the assumption that any problem situation can be described as a mathematical model that supports selection of the best alternative on the basis of some criterion.

The normative theory is grounded on the following premises:

1) first of all, a mathematical model of any situation might be build and it would generate a set of alternatives;

2) secondly, an objective model (a goal model) might be build to estimate the measure of concordance/adequacy of each alternative to the goal.

Initially, the basis for the normative decision theory lies in the classical concept of maximizing the expected utility. According to this concept, an individual is always trying to make the best decision, which corresponds to the maximum expected utility/gain. It is assumed that the decision-maker has enough information about the environment, alternatives and their consequences of his/her choices.

The main elements, which are used in the process of selecting the best alternative from the set of alternatives are formalized in the form of a decision-making task.

The decision-making task is a pair $\langle A, P \rangle$, where: A is a set of alternatives, P is the principle of optimality, which defines the best alternative (that is a model of an objective) [12].

A solution to the task $\langle A, P \rangle$ is a subset $A_{opt} \in A$, which supervenes from the principle of optimality. If the principle of optimality (P) is a set of contradictory criteria, then there is a multi-criteria task.

The main practical result of the research conducted under the normative decision theory is different mathematical models aiming to find a solution to $\langle A, P \rangle$.

A model is an image of (similar to) the original object, which reflects its significant attributes/properties and which is used in the research process (replacing the original object).

A mathematical model of an economic object (process, event) is a description, made by mathematical means, such as expression equations, inequalities and logical relations.

Examples of mathematical economic models can be:

- 1) mathematical economics (production functions and industries' balances);
- 2) models of mathematical programming;
- 3) game theory models;
- 4) statistics, probability and simulation models.

A normative (pragmatic) decision model means the right representation of the decision-making process and its outcome [4]. These models show how to make a decision.

The most common model used for solving multi-criteria problems, is the model based on the total efficiency criteria: all the alternatives payoffs are multiplied by their weight (according to the criteria values).

Step-by-step guide

Step 1. Read the chosen case and try to find out what problem the enterprise faced (this case was used in the previous practice). Formulate the objective (goal) of decision-making.

Run the software for multicriteria decision analysis (hierarchy analysis). It may be the Expert Choice or the Emperor. Create a new model.

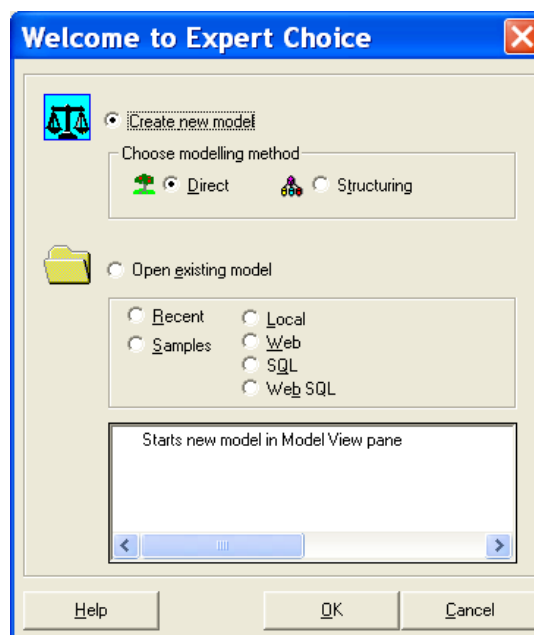


Figure 7. Creation of a new model in the Expert Choice

Enter the model goal, you've formulated.

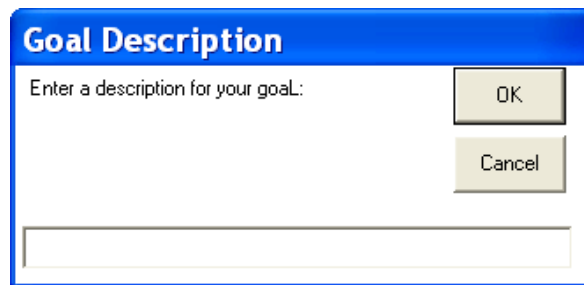


Figure 8. Goal description of a new model

Step 2. Formulate all the actors of this business situation. Ground your choice of people (positions), who are decision-makers for this business situation. Put their positions (they might be put as children of a Goal node) into the model.

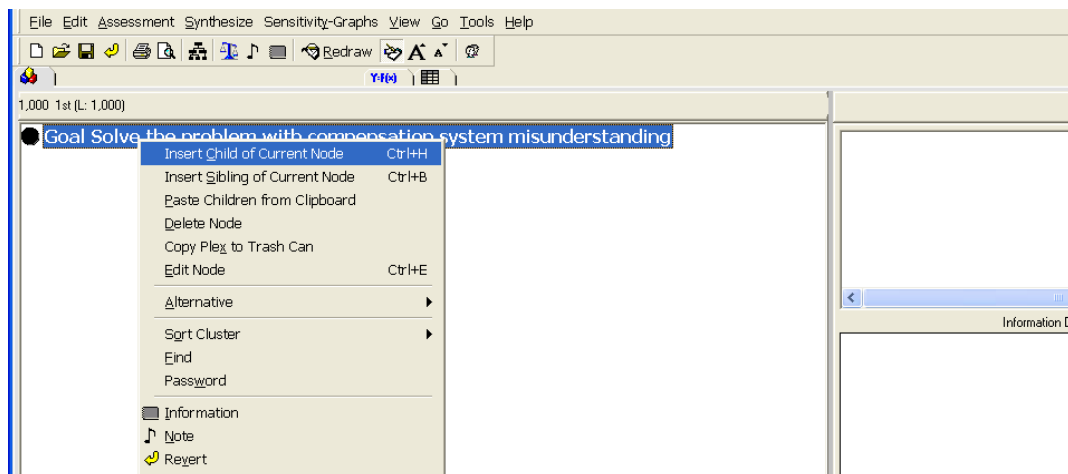


Figure 9. Making a hierarchy structure

All the decision-makers should be put into the hierarchy model in the first level of division.

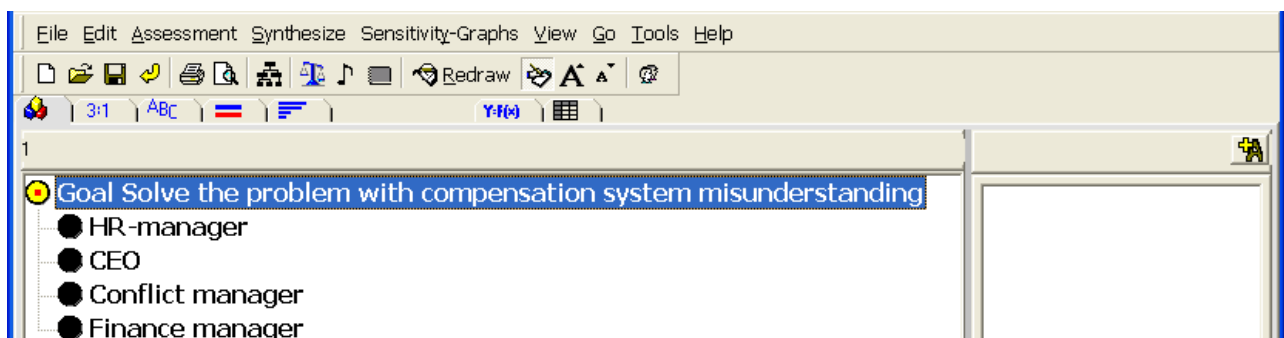


Figure 10. Putting in all decision-makers

Step 3. Put the criteria of all decision makers.

In order to fulfill this step correctly it is obligatory to put yourself in each decision-maker's shoes. Each decision-maker may have his/her individual point of view on the situation of choice because of specific knowledge and experience background, commitments and position. That's why the criteria for evaluation of how alternatives meet the goal might be different for each decision-maker.

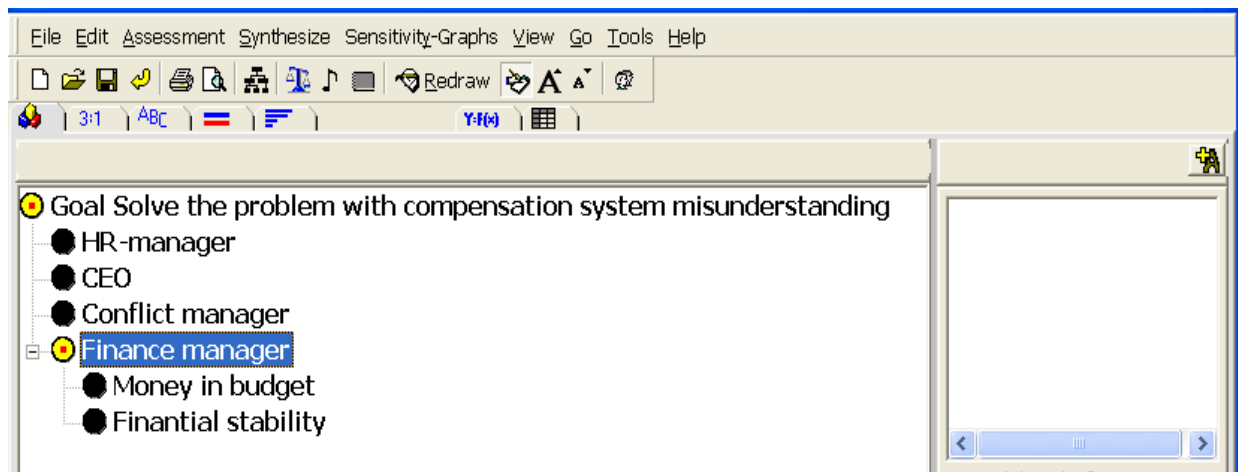


Figure 11. Putting in the criteria of decision-makers

In order to check all the criteria of all the decision-makers we may built a hierarchy model, its first draft. For this purpose we may use the menu **View/Hierarchy view**.

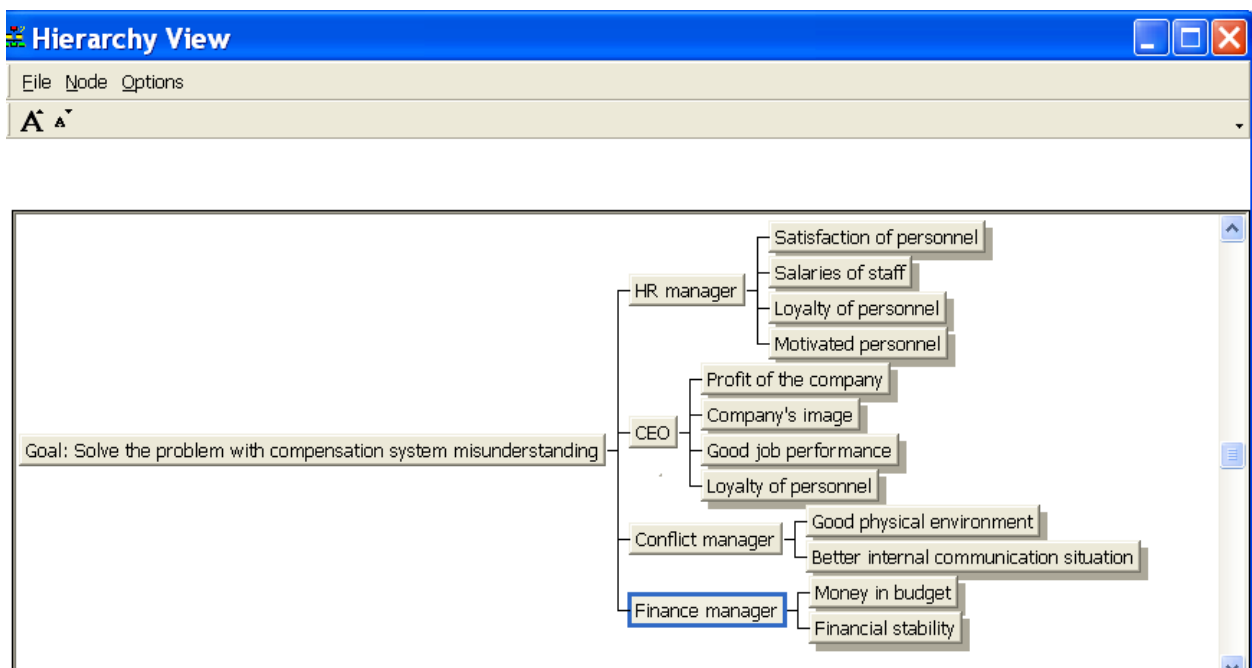


Figure 12. Hierarchy of decision-makers and their criteria

Step 4. Type in all the alternatives.

The last (but not the least) part of the multicriteria decision task is formulating all the alternatives. They should be put to the right part of the window to a special space for alternatives.

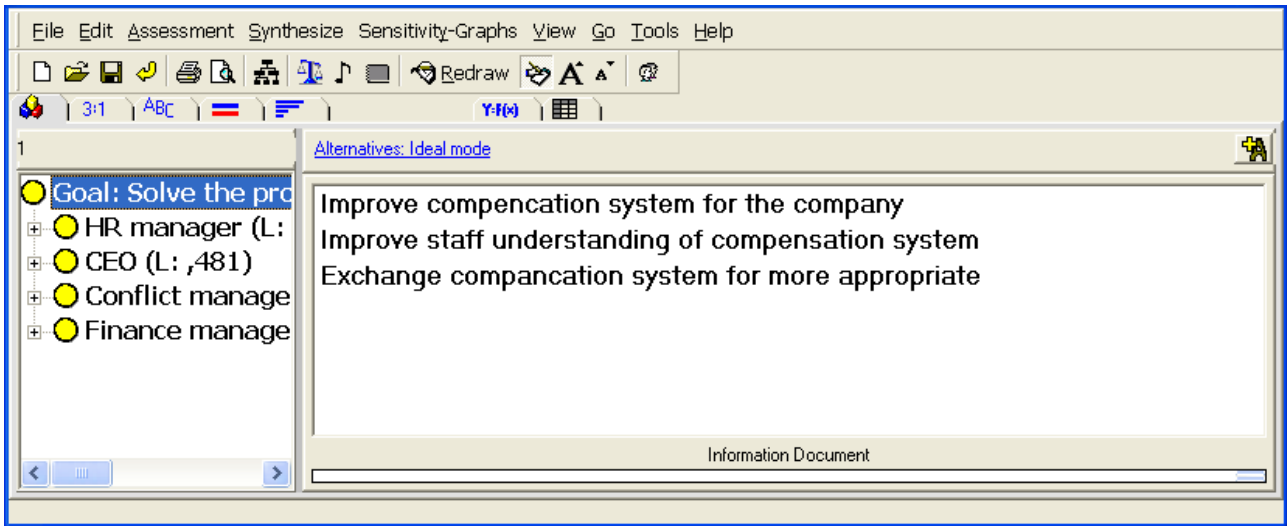


Figure 13. Alternatives of the model

By finishing this step we've provided a multicriteria decision-making task. So that is a perfect time to make an economic description of the model elements in your report.

Step 5. Describe all the preferences in the model.

Preferences assignment may be started with the identification of the weight of each decision-maker's judgment. In our case the most important role is that of the CEO, the least important role belongs to the conflict manager (Fig. 14).

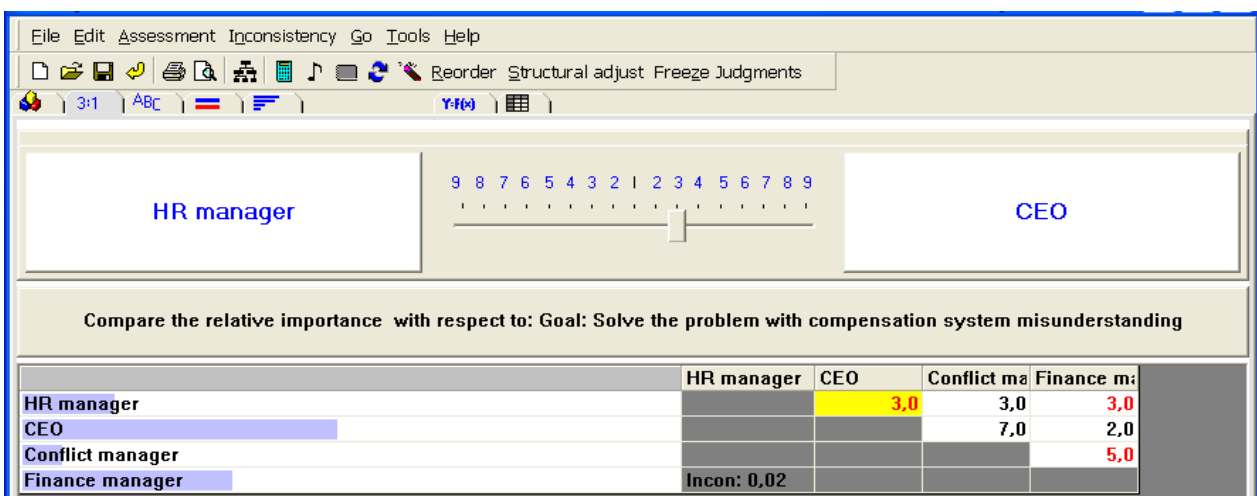


Figure 14. The weight of decision-makers' judgments

Afterwards it is obligatory to put the preferences of decision-makers to their criteria, as it is described in Fig. 15.

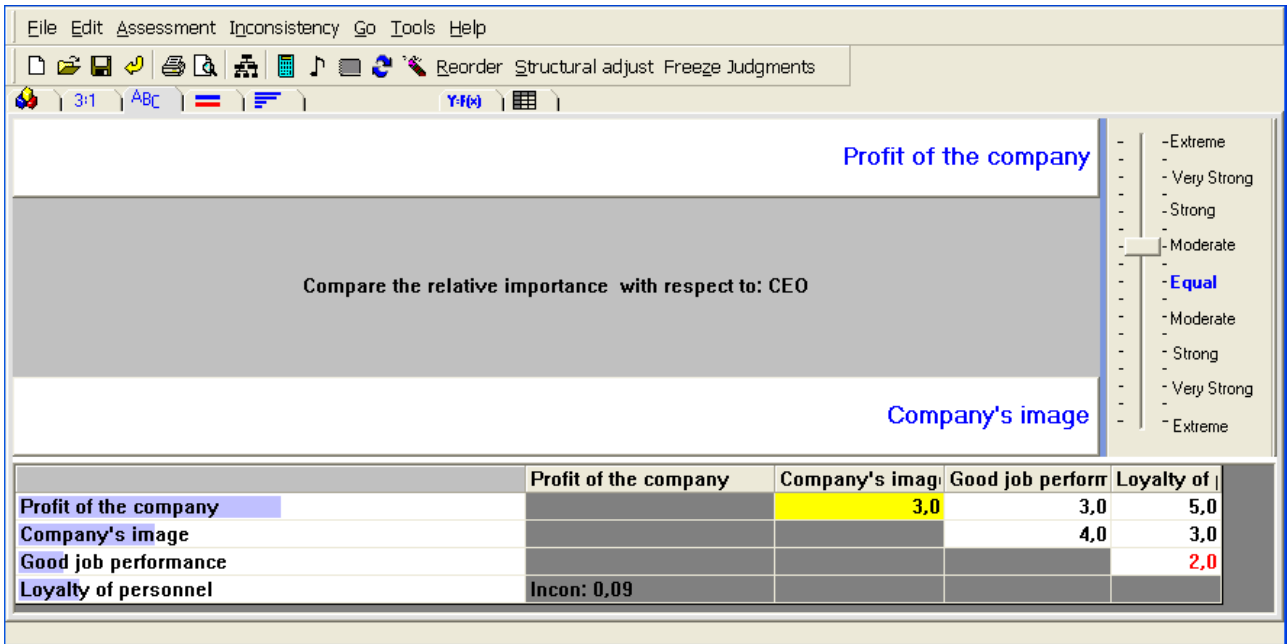


Figure 15. Alternatives of the model

The same way relative preferences of alternatives with respect to all criteria should be assigned.

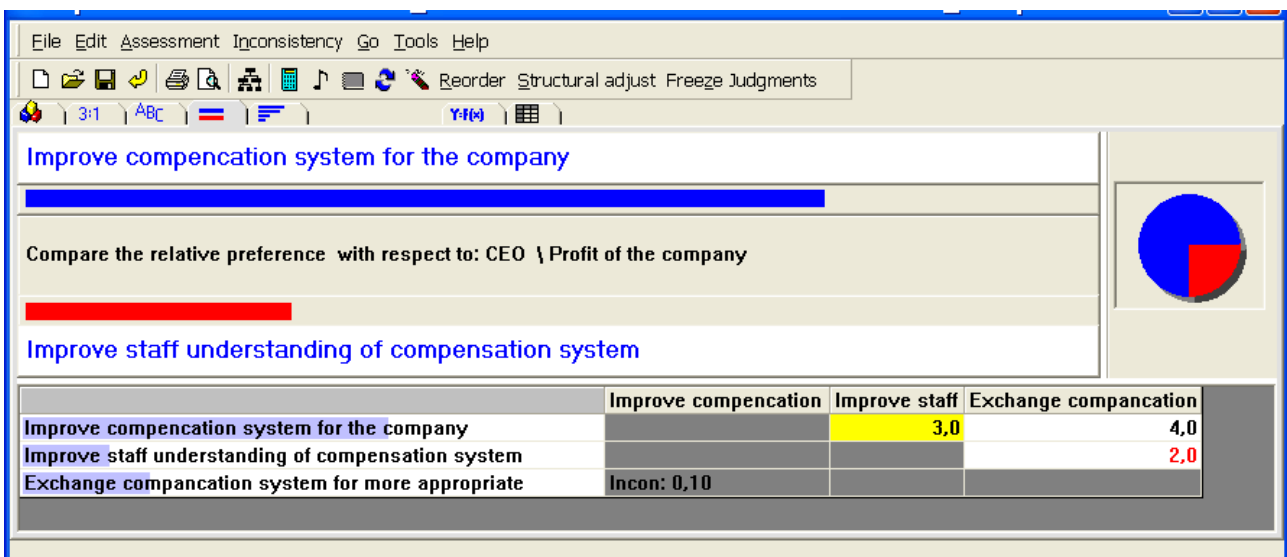


Figure 16. Relative preferences of alternatives on the basis of company's profit criteria from the CEO's point of view

It is easy to see that there are three modes of preference assignment (as shown in Fig. 14 – 16), they are pairwise numerical comparison (Fig. 14), pairwise verbal comparison (Fig. 15) and pairwise graphical comparison (Fig. 16). Any of them might be used in preferences assignment.

Step 6. Calculate the model.

The model may be calculated (or in terms of Expert Choice "synthesized") by using the menu **Synthesize/With respect to the goal**.

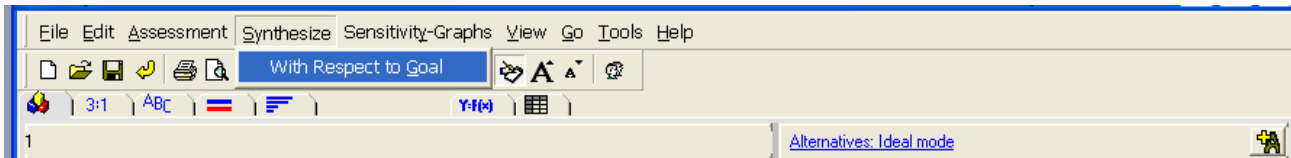


Figure 17. **Model calculation**

The results of the model calculations on a specific tab are called "Priorities derived from pairwise comparisons".

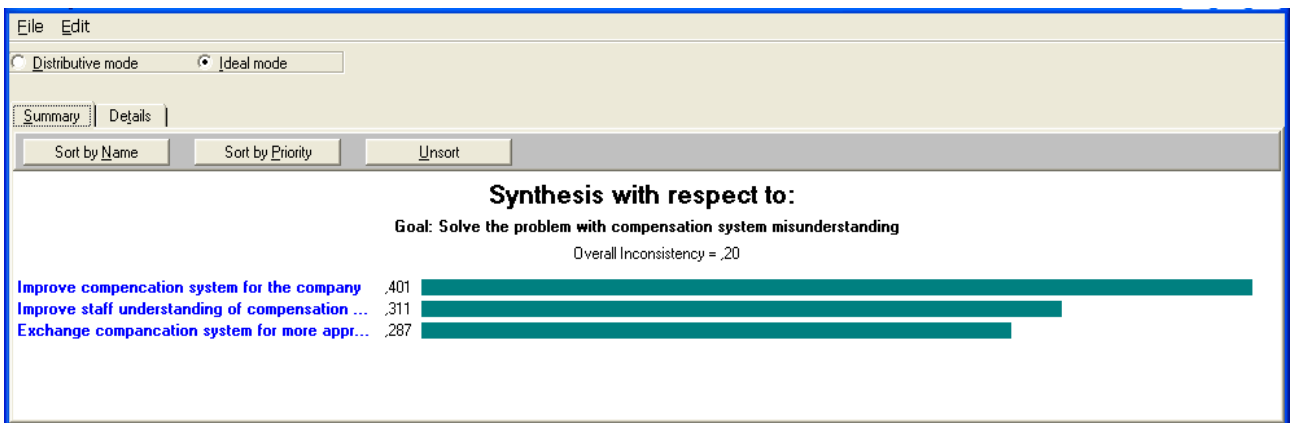


Figure 18. **Graphical representation of the model calculation**

The graph of priorities shows both the priorities themselves and the inconsistency level of the model, which helps to identify the reliability of the model. At the end of model building it is obligatory to analyze the results from the economic background and make a subsequent summary of judgments and results.

The software has a report-making function (the menu **File/Print report**). It provides the researcher with priority graphs, hierarchy, tables of preferences etc. Put it all in your report and explain economic meaning of the model itself and all the results.

Step 7. Make sensitivity analysis.

The last part of multicriteria decision-making is sensitivity analysis which may provide the decision-maker with information of how sensitive the decision is to the changes (even slight changes) in the priorities. Such information may help identify the weak points of the model.

The menu Sensitivity-Graphs provides us with intuitively easy tools of sensitivity check.

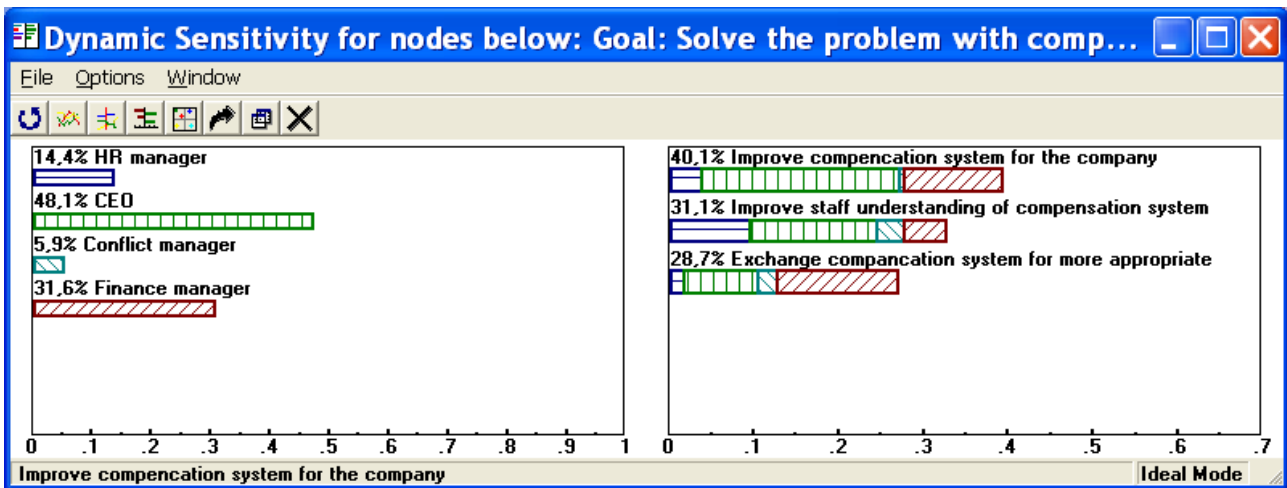


Figure 19. Sensitivity analysis

For example, the Dynamic mode shows the impact of each decision-maker's judgment on the alternative choice. So let's investigate the case of finance manager's importance growth. Even if the CEO and the finance manager are of the same importance, the choice of alternatives doesn't change.

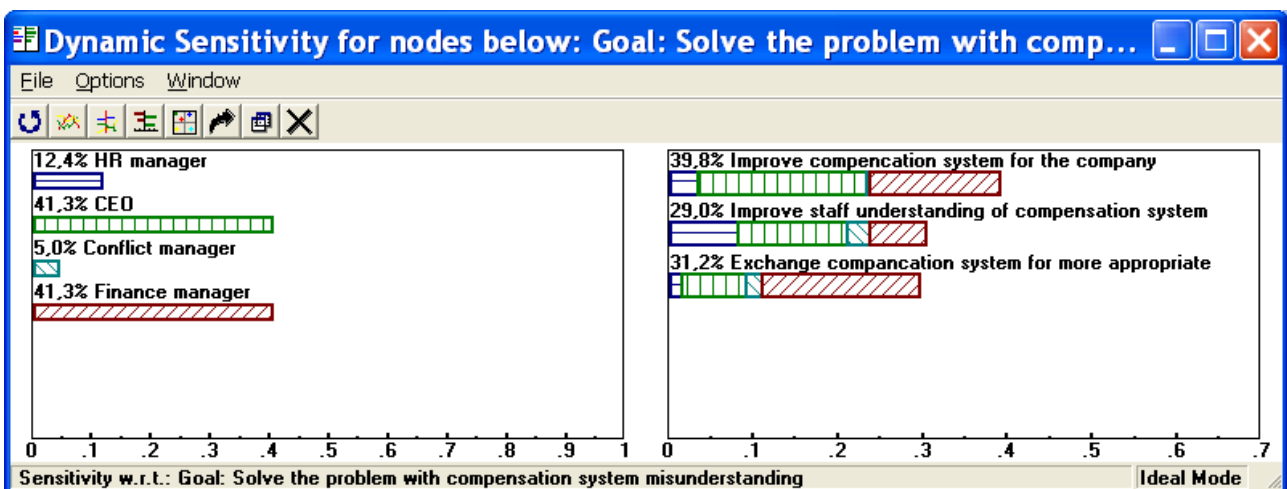


Figure 20. Sensitivity analysis (finance manager's importance growth)

The opposite situation may occur in the case of HR-manager's importance growth. His point of view on the subject is not so aligned with the CEO's point of view. That's why the HR-manager's importance growth leads to the choice of the alternatives significant change.

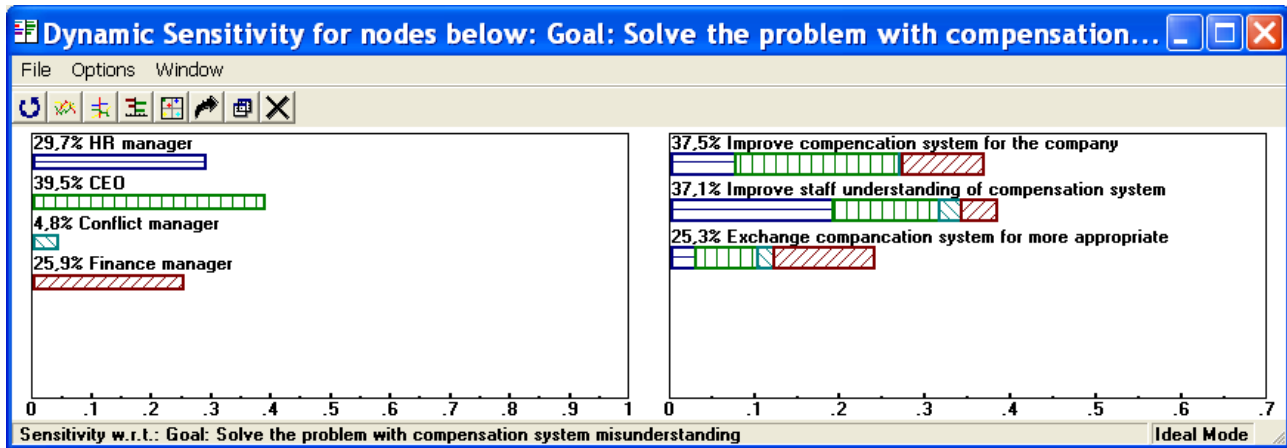


Figure 21. Sensitivity analysis (HR-manager's importance growth)

So the sensitivity analysis provides researchers with additional information about the consequences of change in preferences. Also that is a tool for a wide range of assumptions checks.

Step 8. Make a final report with all the summaries and economic explanation of the analysis results. Submit your practice report, as well as the source Expert Choice file.

Practice 3

Group decision-making. Computer support of group decision-making

Key goals:

- to learn the group decision-making process and its deficiencies;
- to gain the skill in group decision-making processing with the help of the group decision support system (the skill in constructing a questionnaire in the group DSS);
- to acquire the skill in aggregating individual and group expert judgments using the analytic hierarchical process (the skill in processing the questionnaire in the group DSS).

The theoretical background (in short)

In organizations many decisions are made by groups of people, not individuals. If handled in the right way, a decision made by a group can evoke greater commitment than one made by an individual because more people feel the sense of involvement in it. On the other hand, group decisions usually consume more time (and more money) than individual ones, so they need to justify extra costs. In fact it is not possible to generalize about whether individuals or groups are universally better. It depends on the abilities and training of the individuals and groups, and also on the kind of task being tackled [12]. Main decision-making tasks that groups can face are: generating plans; generating ideas; solving problems with correct answers; deciding issues with no identifiably correct answer at the time the decision is being made.

Involving lots of people in the decision-making process provides a greater basis of experience, knowledge, and creative insights. It is intuitively reasonable that the chances of overlooking possible events and possible courses of actions are diminished in group decision-making. The synergy of individuals may make the overall quality of the group decisions greater than the sum of parts [16, 20]. However, if the opinion and values of individuals differ, how should the differences be resolved? Difference of opinion may inhibit the expression of critical ideas, resulting in an incomplete survey of alternative action courses or choices. Techniques to enhance a group's creative potential and interaction such as Delphi, SODA (strategic options development and analysis), and even the brainstorming process are widely used.

In group decision-making there are two main issues in values processing. The first is how to aggregate individual judgments, and the second is how to construct the group choice from individual choices [30]. In reality group decisions should not go by consensus because not all people feel the same about things. The minority can have very strong commitments to a cause and can give rise to disruptions that the majority feels lukewarm about. There is no hiding from this issue in the real world. The reciprocal property plays an important role in combining the judgments of several individuals to obtain a judgment for the group. Judgments must be combined so that the reciprocal of the synthesized judgments must be equal to the syntheses of the reciprocals of these judgments.

It has been proved that the geometric mean is the unique way to do that. If the individuals are experts, they may not wish to combine their judgments but only their final outcome from a hierarchy. In that case one takes the geometric mean of the final outcomes. If the individuals have different priorities of importance their judgments (final outcomes) are raised to the power of their priorities and then the geometric mean is formed.

Using the absolute scale approach of the analytic hierarchical process, it can be shown that due to the fact that now the individual preferences are cardinal rather than ordinal, it is possible to derive a rational group choice satisfying the above four conditions. It is possible because: a) individual priority scales can always be derived from a set of pairwise cardinal preference judgments as long as they form at least a minimal spanning tree in the completely connected graph of the elements being compared; and b) the cardinal preference judgments associated with the group choice belong to an absolute scale that represents the relative intensity of group preferences.

Step-by-step guide

In order to acquire the announced skills some short questionnaires (of websites usability investigation) are given to students. These questionnaires are filled by experts, some judgments are missed, – all they are the parts of real-life website usability investigation.

Step 1. Run the group decision support software (Expert Choice). Make a new model.

Step 2. Put in the goal of the questionnaire as "Analysis of the enterprise website usability".

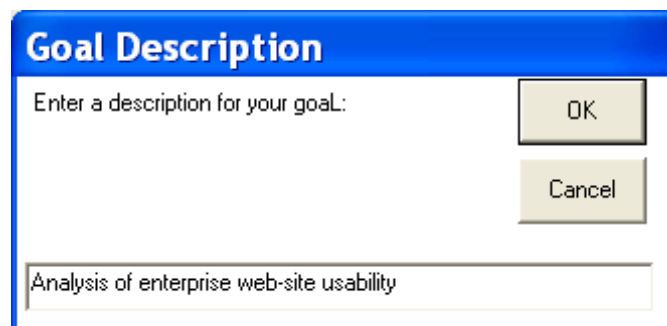


Figure 22. **The goal of questionnaire model**

Step 3. Type in the questions of the questionnaire (as siblings of the goal).

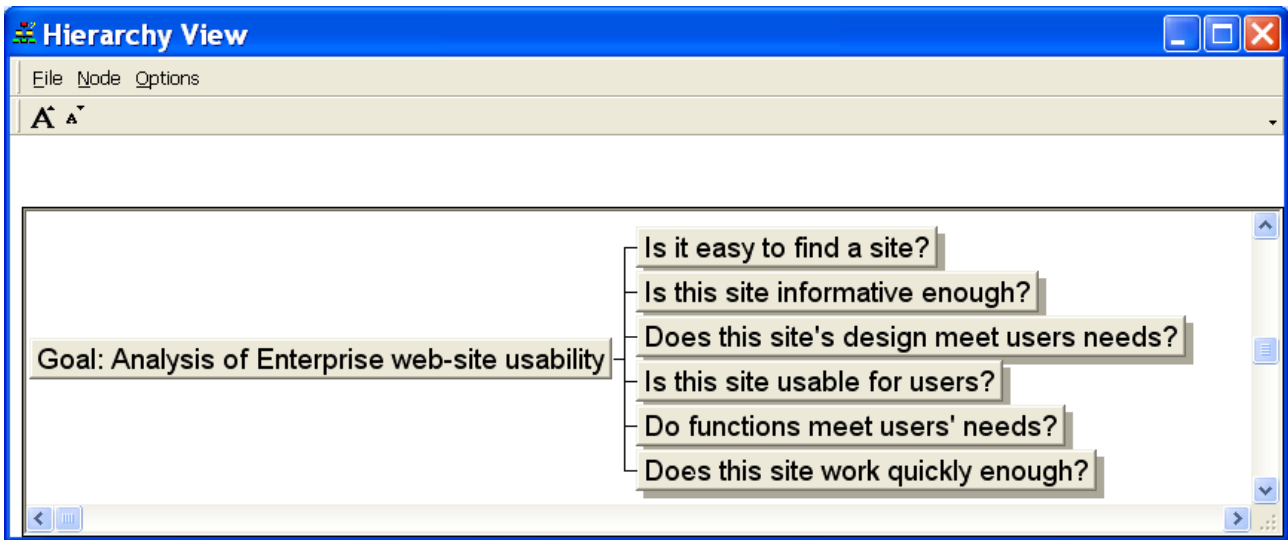


Figure 23. The hierarchy view of the questionnaire model

Step 4. Put in the alternatives of your model.

In the case of the enterprise website usability analysis questionnaire the alternatives are websites which were investigated and estimated by experts. Due to training purposes of this practice there only 4–6 questionnaire sheets are given, where 3–6 enterprises' websites might be mentioned. So put in the alternatives (website's names) to your model.

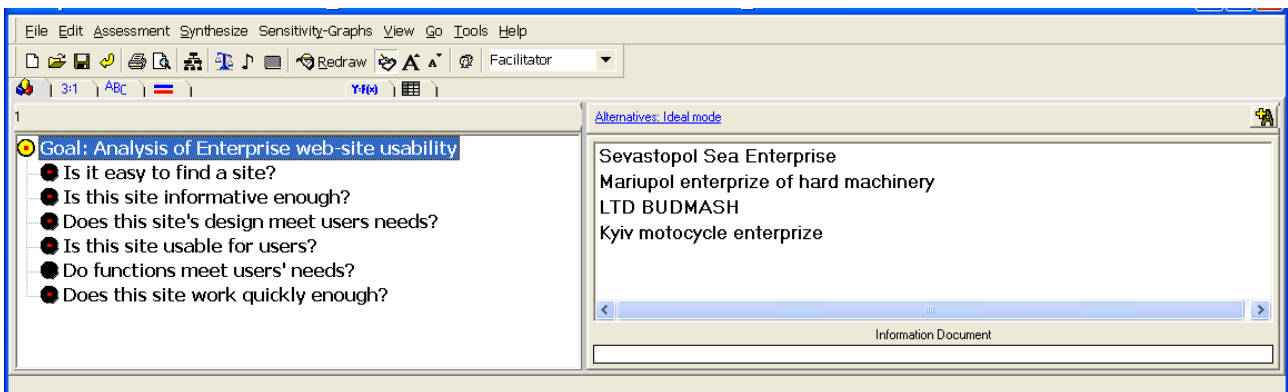


Figure 24. The alternatives of the questionnaire model

Step 5. Define the possible answers grid.

The real-life questionnaires often operate linguistic variables, because they are easier accepted by experts or respondents. So in order the software may process the answers of that kind, it is obligatory to put such answer grid to the model and assign numerical values to it. For that purpose one should fill the tab "Ratings – Define and derive scale priorities for intensities". We've

got five possible answers in the questionnaire: yes, more yes than no, fifty-fifty, more no than yes, no. These answers may correspond to different numerical values (according to study purposes and different distribution bases), but for training purposes let's submit the equal pace between judgments – 0.25.

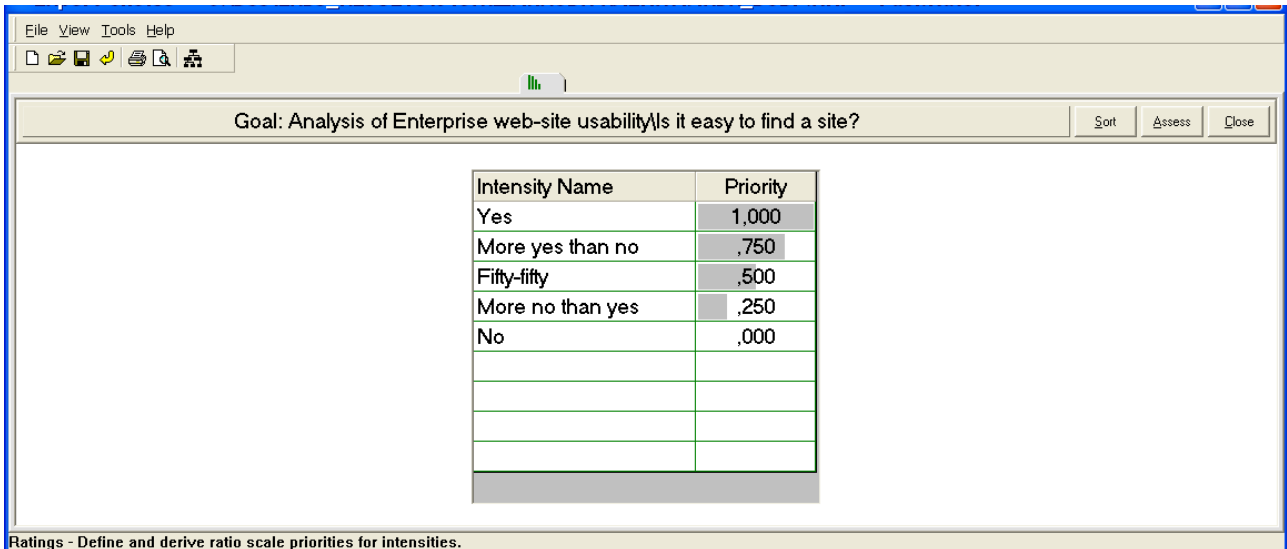


Figure 25. The rating grid

Afterwards one should assign the formed answer grid to questions. The software sustains the possibility of having a different answer grid to different questions, but in our case they are the same. So one has to switch to the "Formulas grid" tab and assign the formed answer grid (ratings type) to every question.

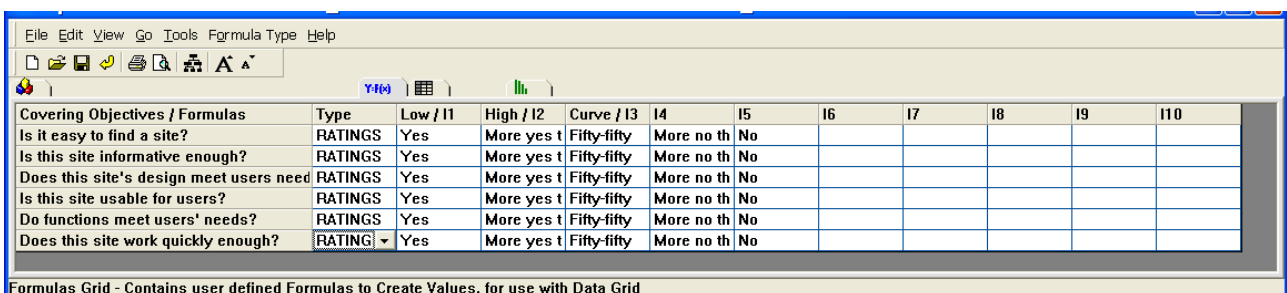


Figure 26. The formula grid

So our model is ready for individual judgment processing. Save it!

Step 6. Add respondents for individual and group expert judgments aggregation using the group DSS.

In order to enrich the model with expert judgments one should create separate "subquestionnaires" for the experts. Use the menu **Go/Participants table**.

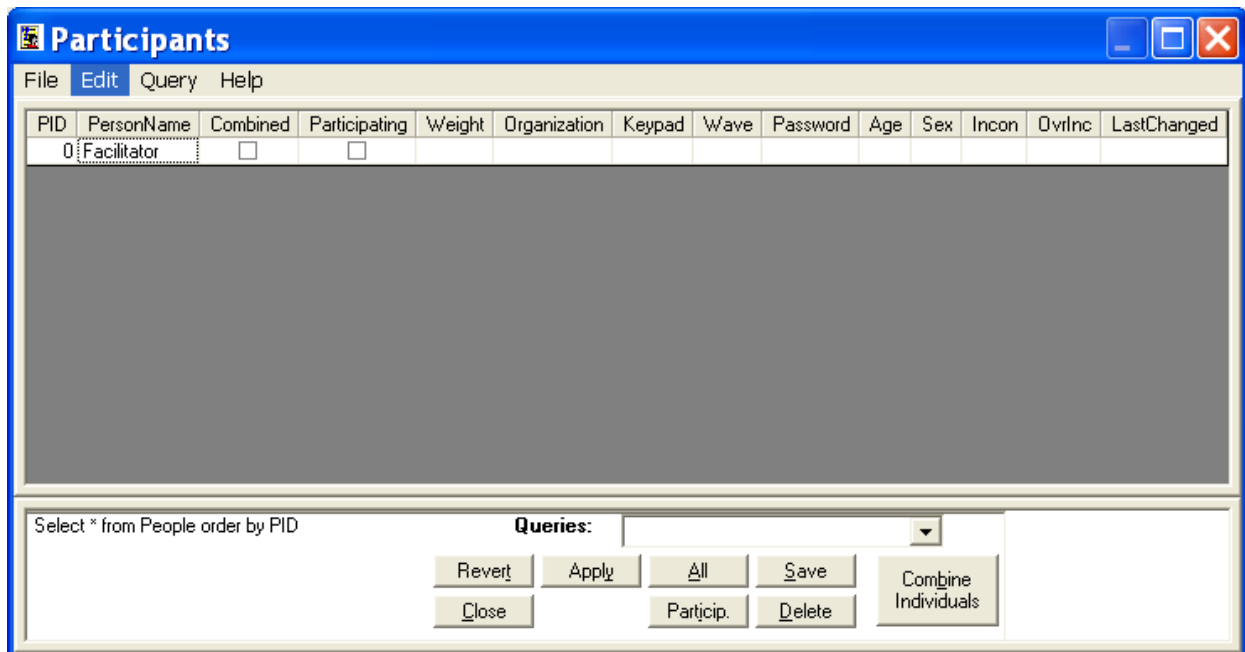


Figure 27. The participants table

In the participants table use the menu Edit/Add N participants. In our case we should add two participants because we are processing four questionnaire sheets filled in by two experts. It is polite to put the names of the experts into the table at once. The role "Combined" will be added automatically, it serves to aggregate judgments.

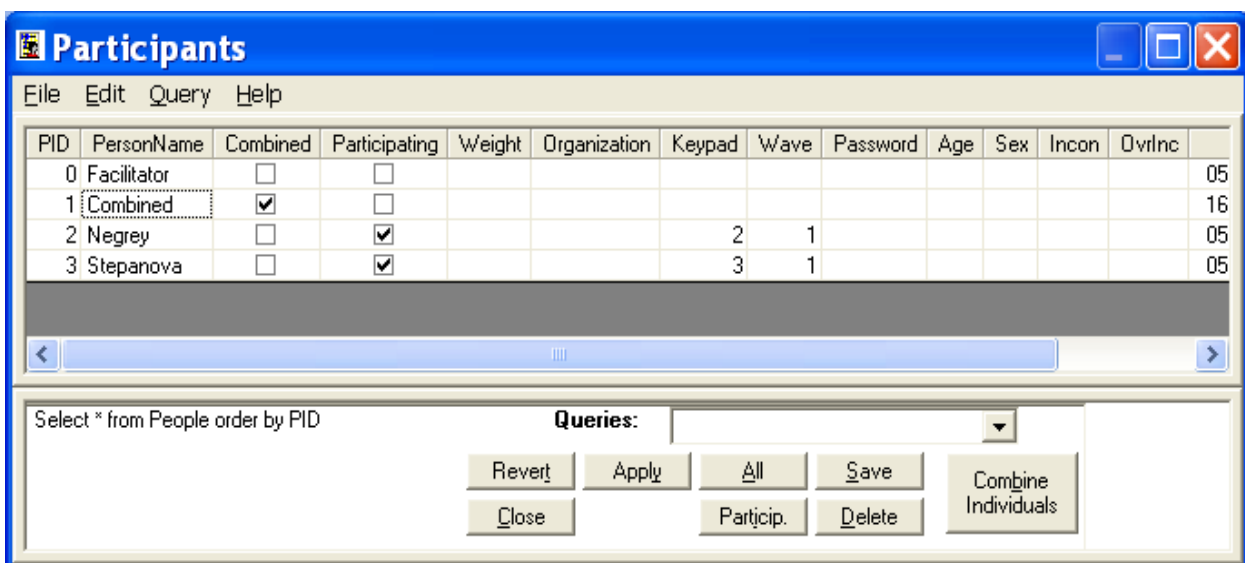


Figure 28. The filled in participants table

At this step the model should be saved, it is ready for preferences to be filled in.

Step 7. Fill in the preferences of experts and combine the results.

In order to fill in the preferences of every expert one has to switch to the chosen participant (expert) and fill in the model with the data from the questionnaire sheets in the Data grid tab. In order to fill the grid tab it is enough just to click the needed answer from the prompt answers at the top.

Yes	More yes than no	Fifty-fifty	More no than yes	No
1 (1,000)	2 (.750)	3 (.500)	4 (.250)	5 (.000)

AID	Alternative	Total	Is it easy to find a site? (L: .167)	Is this site informative enough? (L: .167)	Does this site's design meet users needs? (L: .167)	Is this site usable for users? (L: .167)	Do functions meet users' needs? (L: .167)	Does this site work quickly enough? (L: .167)
A1	Sevastopol Sea	.708	Yes	Fifty-fifty	Fifty-fifty	More yes than no	Fifty-fifty	Yes
A2	Mariupol enterprize of	.000						
A3	LTD BUDMASH	.000						
A4	Kyiv motorcycle	.625	Yes	Fifty-fifty	More yes than no	More no than yes	More no than yes	Yes

Figure 29. Filling in one expert data grid

After filling in the data from all the questionnaire sheets one should switch to the Combined role, where all the data are preformed in one view.

Yes	More yes than no	Fifty-fifty	More no than yes	No
1 (1,000)	2 (.750)	3 (.500)	4 (.250)	5 (.000)

AID	Alternative	Total	Is it easy to find a site? (L: .167)	Is this site informative enough? (L: .167)	Does this site's design meet users needs? (L: .167)	Is this site usable for users? (L: .167)	Do functions meet users' needs? (L: .167)	Does this site work quickly enough? (L: .167)
A1	Sevastopol Sea	.708	1,000	.500	.500	.750	.500	1,000
A2	Mariupol enterprize of	.833	1,000	1,000	.750	.500	.750	1,000
A3	LTD BUDMASH	.792	1,000	.750	.500	.750	.750	1,000
A4	Kyiv motorcycle	.625	1,000	.500	.750	.250	.250	1,000

Figure 30. The aggregated results of the website usability analysis

As a result all the weak points of the websites are well seen, the best alternative (the best website of the enterprise) is evident.

For visible indication of the best alternative one can use the **Plot alternatives** command.

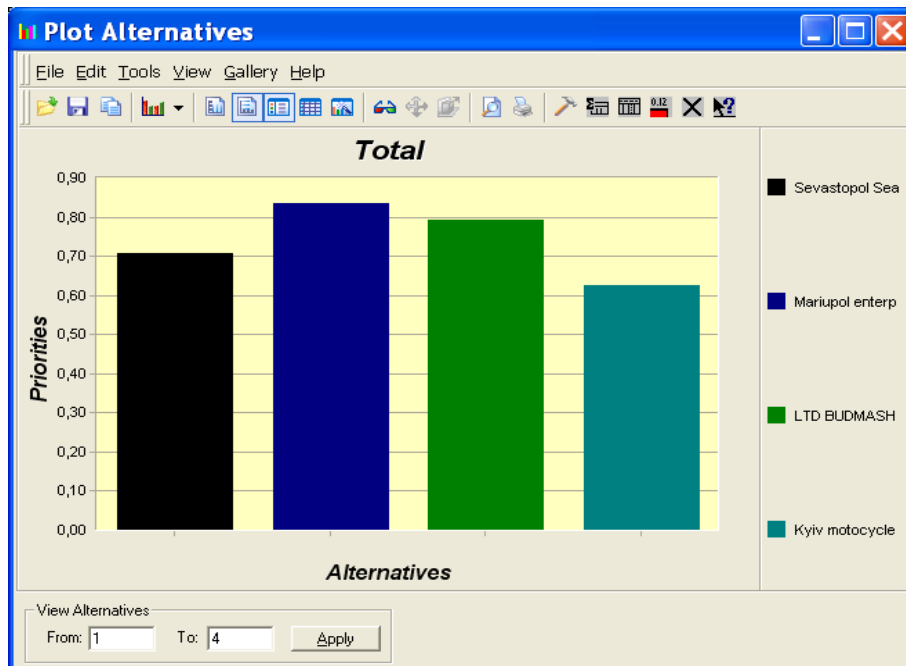


Figure 31. The website usability analysis results

Step 8. Make a final report with all the summaries and economic explanation of analysis results. Submit your practice report, as well as the source Expert Choice file.

Practice 4

Computer support of group decision-making on individual data

Key goals:

to gain the skill in the group decision-making processing with the help of the group decision support system (the skill in constructing questionnaires in the group DSS) on individual data;

to acquire the skill in making aggregate individual and group expert judgments using the analytic hierarchical process (the skill in processing questionnaires in the group DSS) on individual data.

Step-by-step guide

Step 1. Make a questionnaire of your own. Define the goal of the analysis (Why is it necessary to fill in this questionnaire?), what new knowledge you are going to get. Test your questionnaire on someone you

trust, improve your questionnaire in the case of a negative feedback. Define your experts and poll the assembly.

Step 2. Run the group decision support software (Expert Choice). Make a new model.

Step 3. Put in the goal of your questionnaire. Type in the questions of the questionnaire (as siblings of the goal).

Step 4. Put in the alternatives of your model.

Step 5. Define a possible answer grid.

Step 6. Add respondents for individual and group expert judgments aggregation using the group DSS.

At this step the model should be saved, it be ready for preferences to be filled in.

Step 7. Fill in the preferences of the experts and combine the results.

Step 8. Make a final report with all the summaries and economic explanation of the analysis results. Submit your practice report, as well as the source Expert Choice file.

Module 2

Systems and technologies of management decision support

Practice 5

Data search for industry analysis

Key goals:

to learn data types;

to gain the skills in working with potential sources for raw data;

to gain primary skills in preparing and preprocessing data;

to make analysis of the chosen industry.

The theoretical background (in short)

One of the most important objects of the enterprise external environment analysis is industry analysis (industry to which the enterprise belongs to). On the basis of industry state it is possible to draw conclusions about market perspectives, resources accessibility for development, level of competition and market concentration. The economic activity types are grouped in some industries in order to aggregate data on economically (or technically) close business activities and therefore trace macroeconomic changes.

Grouping of business activities in Ukraine is based on the Classification of Economic Activities "КВЕД" (Класифікація видів економічної діяльності). At the moment the latest (and actual) edition is КВЕД of 2010 (the National classifier of Ukraine ДК0009:2010 "Класифікація видів економічної діяльності" <http://kved.ukrstat.gov.ua/>). Till January 1, 2012 КВЕД-2005 was used (the one revised in 2005). At the international level the NACE classifier in two versions is widely used – rev. 2.0 (the recent one, used since 2008) and rev. 1.1 (out-of-date). The NACE rev. 2.0 almost exactly corresponds to the Ukrainian classifier КВЕД-2010 in economic activities grouping. It is available for download on the Eurostat website.

In order to analyze any industry, common indexes (applied to describe any industry) and specific indexes (specific to particular industries) are used.

Common indexes include:

1) indicators of production and sales of products (production volume, sales, imports and exports volume in physical terms and monetary terms);

2) financial indicators (gross income, operating surplus volume, net profit, profitability);

3) social characteristics of an industry (the number of staff employed and its dynamics, the average wage level, the average compensation costs, staff turnover, personnel training, etc.);

4) indicators of innovation orientation and investment attractiveness (investment volume and sources of these investments, investments objects: equipment, new technology, depreciation of capital, the percentage of innovative products, R&D personnel in industry, etc.);

5) ICT usage (enterprises using Internet in order to sale products, get orders, for supply chain and inventory tracking and tracing, enterprises that use the Internet for banking and financial services, enterprises with website providing product catalogues or price lists).

Specific industry indicators reflect the peculiarity of their functioning. For example, the number of mobile operators subscribers shows the level of market penetration of this service; the length of railways or roads paved describes the infrastructural capability to implement the transport and logistics processes, etc.

ICT usage in accounting and statistical information aggregation, as well as promotion of the idea of openness in developed societies have led to the availability of free access to such international statistical databases:

1) the European Commission of Statistics (<http://epp.eurostat.ec.europa.eu/>);

2) the Organization for Economic Co-operation and Development (<http://stats.oecd.org/>);

3) the United Nations (<http://www.un.org/en/databases/>);

4) the World Bank (<http://data.worldbank.org/>).

National statistical committees websites of world economic leaders are also available, such as the USA (<http://www.usa.gov/>), Australia (<http://abs.gov.au/>), China (<http://www.stats.gov.cn>), Russia (<http://www.gks.ru/>).

The nice thing about working with the majority of these global statistical databases is the ability to provide information under the user requests (such customization is possible in varying degrees, but it gives access to highly specialized data). Also these databases allow users to export the data in a format suitable for analysis, which significantly reduces the amount of manual labor of users.

The main source of statistical data in Ukraine is the State Statistics Service of Ukraine (the Ukrainian Statistics Committee or so called Ukrstat). Its data is available in the electronic (online <http://www.ukrstat.gov.ua/>) and printed form. The main drawback of its data is the fact that the content from these sources overlaps but is not identical. Not all papers can be found in the electronic form, while "Експрес-випуски" (Express Editions) are not represented in the printed form at all (they are only available in Ukrainian).

The data are presented in three main section on the Ukrstat website:

1) "Publications" ("Публикации", this section allows users to download an electronic analog of the printed edition, the data in such publications cannot be resaved in the formats appropriate for automated analysis);

2) "Express Editions" ("Експрес-випуски" contain operational statistics, they are often presented in the context of months and quarters aggregation. This section is not presented in English);

3) "Statistical Information" (This section presents data tables on a web page, so it allows users at least partially transfer data to other programs and not retype them. The main drawback of the section is that provided information is cut).

One of the important aspects of industry analysis is collection and analysis of key market players, the dynamics of their functioning and development. Enterprises' websites study may be used as sources of such information, but nowadays the penetration of IT technologies in industries of Ukraine is uneven, and the information provided on these websites is often cut and outdated.

Due to the mentioned website features of many Ukrainian enterprises, it is effective to use the official website of the Agency of Stock Market Infrastructure Development of Ukraine (<http://www.smida.gov.ua/about>) for gathering information about public enterprises. This website uses the publicly

available database of the Stock Market Commission. The following data is available (which is vital for key players analysis):

- 1) the balance sheet statement (форма 1 "Баланс предприятия");
- 2) the financial results statement (форма 2 "Финансовые результаты");
- 3) personnel;
- 4) wage and earnings.

Also there might be some information about the production volume, the net cost, etc., though not all the issuers publish such data.

For efficient use of the database it is obligatory to register and keep in mind this website rule: viewing more than one report for a minute is forbidden. Without registration companies' reporting statements are not available at all. The main drawback of this website is its availability for Ukrainian issuers only.

The core sources of information about the industry position are analytic reports of international consulting agencies. Some of them are available for free. Such reports may describe not only the recent information about the researched industries, but also include some specific industry indexes, which were generated (or gathered) by such agencies. The analysis of talent attraction and compensation in communication industry made by PriceWaterCoopers might be mentioned as an example.

Worldwide famous consulting agencies:

- 1) PriceWaterCoopers (www.pwc.com);
- 2) KPMG International Cooperative (www.kpmg.com/global/en/pages/default.aspx);
- 3) McKinsey & Company (www.mckinsey.com/);
- 4) The Boston Consulting Group, Inc. (www.bcg.com/);
- 5) Bain & Company (www.bain.com/);
- 6) Booz & Company (www.booz.com/);
- 7) Ernst & Young LLP (www.ey.com/) and others.

In order to produce an industry (analytic) report of high quality one should do the following:

- 1) define industry codes according to statistic classifiers NACE rev. 2.0 and КВЕД-2010;
- 2) determine specific features of the researched industry, specify the list of the researched indexes and the time period (if needed);
- 3) search and customize the needed data from international statistic databases;
- 4) collect the available data from national statistical databases, such as Russian and Ukrainian;

- 5) search ready-made analytic reports of the industry;
- 6) make primary analysis of the gathered information, to find world and Ukrainian market players;
- 7) search the major players' statements and reports on the website of the Agency of Stock Market Infrastructure Development of Ukraine, as well as in other open sources;
- 8) summarize the collected information and organize it in the form of a report (or presentation).

Step-by-step guide (in short)

Step 1. In order to investigate the dynamics of the chosen industry (in the world or European market) find the data in open sources (years 2007 – 2015). Such government sources may be the European Commission of Statistics (www.epp.eurostat.ec.europa.eu), EU KLEMS Project (www.euklems.net), etc. Private sector sources may be used: different consulting firms like PwC (www.pwc.com), KPMG (www.kpmg.com/global/en/pages/default.aspx), etc. These sources may help in finding out the key trends in the industry and their volume in the European market (the number of employees, the volume of investments etc. are also available). The better data is found the better industry analysis might be done. Save all the raw data in your folder. Include the summary about the key industry trends in your report.

Step 2. Analyze the dynamics of the industry in the Ukrainian market. The main government source is the State Statistics Service of Ukraine (www.ukrstat.gov.ua). Other nongovernment sources available might be used.

Step 3. Download the financial data of prominent Ukrainian enterprises (they should belong to the chosen industry). Among the saved data there should be: the balance ("Форма 1: Баланс підприємства") and the statement of financial results ("Форма 2: Фінансові результати"); the annual number of employees and the annual wage fund. Save all your results in an Excel file.

Step 4. Compile the results of the analysis (steps 1, 2, 3) into one holistic report of the chosen industry. Be attentive in putting in all the references, which were used!

Step 5. Make a final report with all the summaries and economic explanation of the analysis results. Submit your practice report, as well as **ALL** the source files.

Step-by-step guide (An example of *my* analysis)

Step 0. The industry I've chosen for investigation is publishing. As far as I don't remember the code of the industry (or the main products of that industry) I should find it in statistical classifiers like "КБЕД" or NACE rev. 2. These codes may help in the search of the industry specific data.

In my case the industry is called "Printing and reproduction of recorded media" (division 18).

Division 18	Printing and reproduction of recorded media
18.11	Printing of newspapers
18.12	Other printing
18.13	Pre-press and pre-media services
18.14	Binding and related services
18.20	Reproduction of recorded media

Figure 32. Printing industry codes according to NACE rev. 2.0

According to Ukrainian statistics the publishing industry code is 18 (according to "КБЕД-2005", the 2005 year version) or 22 (in "КБЕД-2010", the 2010 year version).

PRODCOM	Description	External trade nomenclature reference for 2012 (HS/CN)	Volume Physic.unit	P	Reference to Notes
NACE 18.11 : Printing of newspapers					
CPA 18.11.10 : Newspaper printing services					
18.11.10.00	Printed newspapers, journals and periodicals, appearing at least four times a week			S	
NACE 18.12 : Other printing					
CPA 18.12.11 : Printing services for postage stamps, taxation stamps, documents of titles, smart cards, cheques and other security papers and the like					
18.12.11.00	Printed new stamps, stamp-impressed paper, cheque forms, banknotes, etc			S	P
CPA 18.12.12 : Printing services for advertising catalogues, prospects, posters and other printed advertising					
18.12.12.30	Printed commercial catalogues			S	P
18.12.12.50	Printed trade advertising material (excluding commercial catalogues)			S	P
CPA 18.12.13 : Printing services for journals and periodicals, appearing less than four times a week					

Figure 33. Printing industry codes according to NACE rev. 2 (by products)

Step 1. In order to investigate the European dynamics of the printing industry, it is obligatory to start with the European Statistics Office website (<http://www.epp.eurostat.ec.europa.eu>). It provides researchers with raw data, reports and reviews (mostly) for free. The data from the Eurostat statistic databases may be extracted in a wide range of formats, so it is very convenient for making analysis and reports of different kinds.

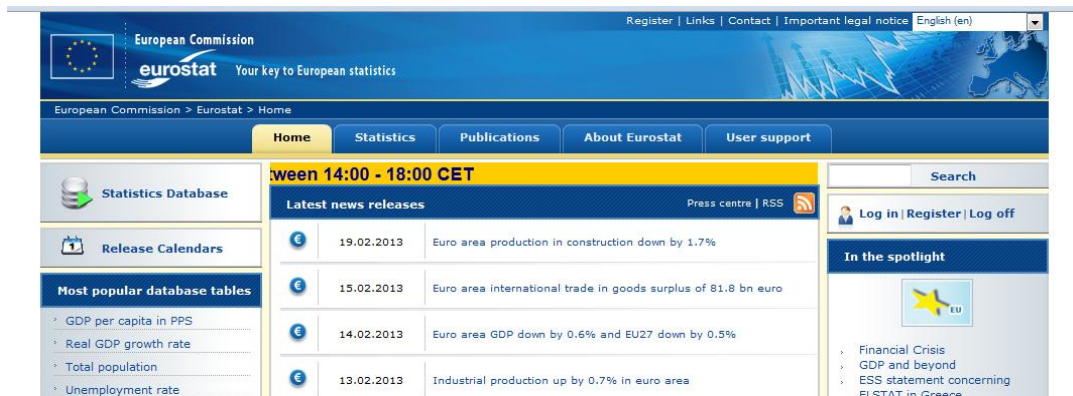


Figure 34. The main page of the European Commission of Statistics website

The easiest start of the search on a website is using the search function. Not all the found results suited the investigation purposes, but some reports were very valuable. It was the industry analytical report of the year 2010.



Figure 35. The analytical report from the Eurostat

The quality of this research paper is high. It includes data about the industry turnover, the number of employees, the value added, the gross operating rate, the investment rate, etc.

2006, around EUR 8 000 per employee higher than the non-financial business economy average. The relatively high level of average personnel costs for this sector was observed in most of the Member States with data available.

recorded media sector in 2006, both accounting for around one fifth of the EU-27 value added (EUR 22.6 billion in the United Kingdom and EUR 18.9 billion in Germany). In employment terms, Germany's share was slightly larger, while that of the United Kingdom was significantly smaller.

Table 1: Structural profile of publishing, printing and reproduction of recorded media, EU-27, 2006

	Number of enterprises	Number of persons empl. (2)	Turnover (2)	Value added (2)	Apparent labour product. (EUR 1 000/ pers. emp.)	Gross operating rate (3)	Invest. rate (2)
	(1 000)		(EUR million)			(%)	
Publishing, printing, reproduction of recorded media (1)	219.9	1 820	260 000	97 000	53.1	13.0	10.0
Publishing	81.0	800	130 000	49 000	60.0	13.1	5.5
Printing and service activities related to printing	132.8	961	105 942	41 647	43.3	13.4	15.7
Reproduction of recorded media	6.2	36	18 713	5 685	156.4	23.9	8.5

(1) Including estimates.
 (2) Publishing, estimates.
 (3) Publishing, printing and the reproduction of recorded media, and publishing, 2005.
 Source: Eurostat ([sbs_na_2a_dade](#))

Figure 36. Data sets in the available analytical report

Also I download statistic yearbooks (years 2007 – 2014) to make my analysis more up-to-date. Some data tables are available in the .xls format (not .pdf), so they may be easier used in calculations.

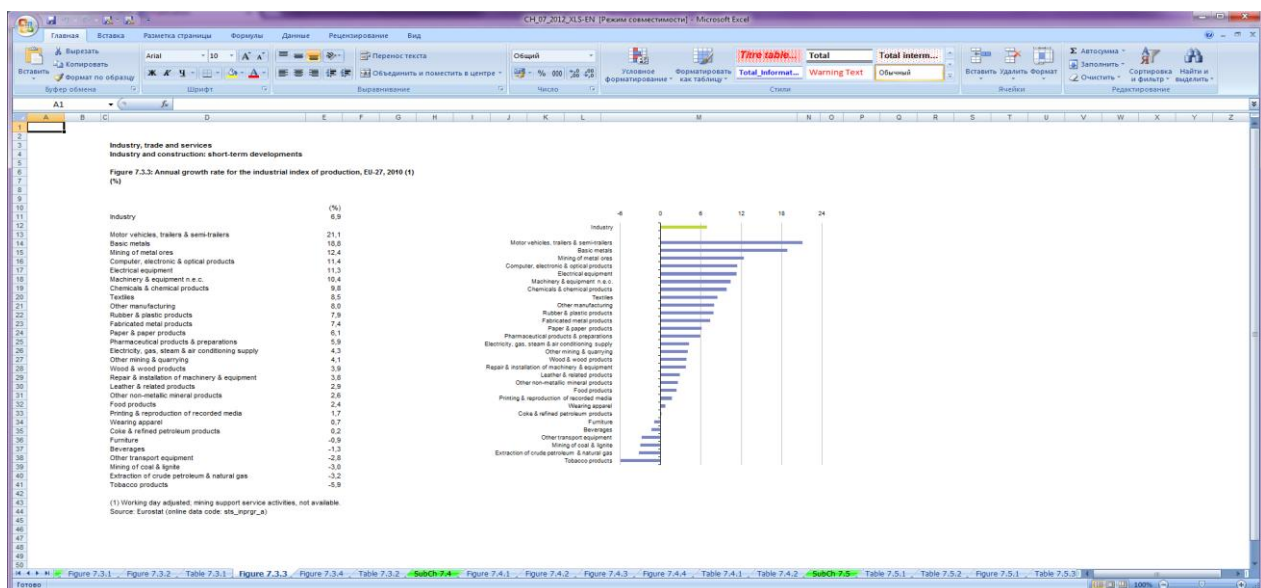


Figure 37. One of the spreadsheets from the yearbook of 2011

For more specific industry analysis the database may be used. These tables provide users with customized data extraction.

	2010	2011	2012	2013
EU	75,000	75,000	75,000	75,000
EU (excl. UK)	75,011	74,712	75,000	75,000
EU (excl. UK) (27)	1,507	1,538	1,625	1,625
EU (excl. UK) (27)	1,064	1,049	1,075	1,075
EU (excl. UK) (27) (excl. Ireland)	5,443	5,245	5,166	5,166
EU (excl. UK) (27) (excl. Ireland) (excl. Malta)	1,095	1,089	1,046	1,046
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus)	6,225	6,007	6,653	6,653
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg)	276	278	311	311
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands)	436	421	430	430
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium)	2,588	2,588	2,588	2,588
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria)	6,951	6,990	7,091	7,091
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece)	9,193	9,318	8,986	8,986
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal)	1,852	1,860	1,844	1,844
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal) (excl. France)	5,823	5,759	5,491	5,491
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal) (excl. France) (excl. Spain)	127	127	120	120
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal) (excl. France) (excl. Spain) (excl. Italy)	424	378	424	424
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal) (excl. France) (excl. Spain) (excl. Italy) (excl. Germany)	517	512	503	503
EU (excl. UK) (27) (excl. Ireland) (excl. Malta) (excl. Cyprus) (excl. Luxembourg) (excl. Netherlands) (excl. Belgium) (excl. Austria) (excl. Greece) (excl. Portugal) (excl. France) (excl. Spain) (excl. Italy) (excl. Germany) (excl. Denmark)	88	88	85	85

Figure 38. Customized data tables from the Eurostat database

For analysis of a specific (narrow) niche the data of each product production volume by countries may be downloaded in the .xls format by product codes (we use info from step 0), so-called PRODCOM annual data.

European Commission > Eurostat > Prodcom - statistics by product > Data > Excel files (NACE Rev. 2)

Home | **Statistics** | Publications | About Eurostat | User support

Prodcom - statistics by product

- Introduction
 - Europroms
 - Steel Data
- Data
 - [Excel files \(NACE Rev. 2\)](#)
 - Excel files (NACE Rev. 1.1)
 - Database

Excel files - NACE Rev. 2

The Excel files on this page contain Prodcom data based on NACE Rev. 2. For 2008 onwards the files contain the original data based on NACE Rev. 2 as supplied by the reporting countries. For 1995 to 2007 the files contain data that has been converted where possible from the data based on NACE Rev. 1.1.

These tables only show the production data. To use the database that contains both production and related trade data, please click on "Database" in the box to the left.

- Prodcom Annual Data 2011 (updated 06/02/2013)
- Prodcom Annual Data 2010 (updated 06/02/2013)
- Prodcom Annual Data 2009 (updated 06/02/2013)
- Prodcom Annual Data 2008 (updated 06/02/2013)
- Prodcom Annual Data 2007 (updated 13/02/2012)
- Prodcom Annual Data 2006 (updated 13/02/2012)
- Prodcom Annual Data 2005 (updated 09/11/2009)

Links

This document describes how to use the Excel files

Figure 39. PRODCOM annual data

These Excel files consist of data by product, it gives an opportunity to analyze market niches and market on the whole (by adding all products).

Afterwards a search on consulting agencies was made. The report on newspaper digital circulation was found on the PriceWatersCoopers' site.



Figure 40. Reports on the PwC site (www.pwc.com)

Step 2. Analyze the dynamics of the industry in the Ukrainian market. The main government source is the State Statistics Service of Ukraine website (www.ukrstat.gov.ua). There the Ukrainian statistics on the industry was found and downloaded as reports in the .pdf format (unfortunately, data customization services aren't provided).

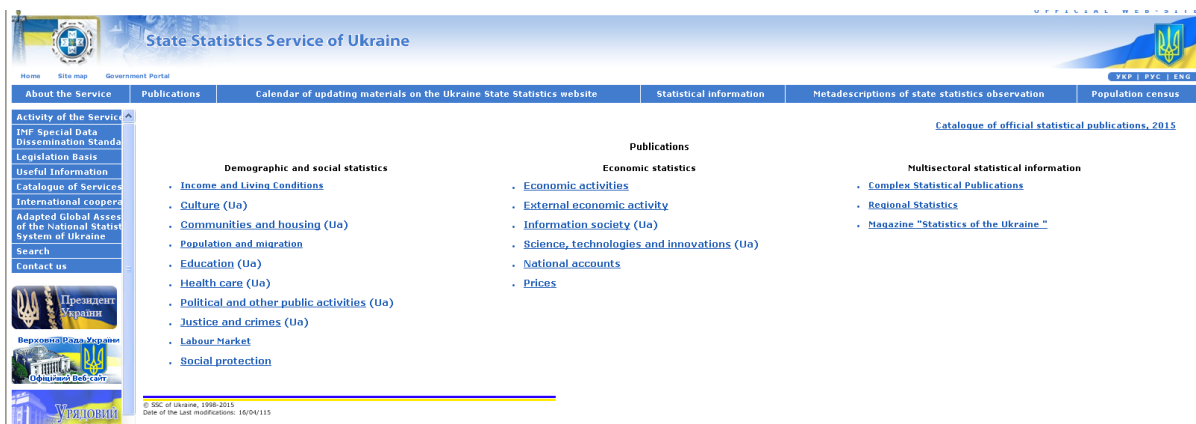


Figure 41. The State Statistics Service of Ukraine website (www.ukrstat.gov.ua)

Step 3. In order to download the financial data of the industry key players, we provided primary analysis. Some of the key players have their websites, so all of them were investigated in search of the needed information. Some of the key players are stock companies, so their financial statistics are aggregated on the official website of the Agency of Stock Market Infrastructure

Development of Ukraine (<http://www.smida.gov.ua/about>). It gathers information about public enterprises. This website uses publicly available database of the Stock Market Commission.

The screenshot shows the search interface on the Smida website. At the top, there is a navigation bar with links: Про агентство, Новини, Заходи, Бази даних, Послуги, Web-сайти, Підтримка, and a search field labeled 'Введіть ЄДРПОУ'. Below this is the search section 'Пошук учасників' with a search input field containing 'глобус' and a dropdown menu for 'Область' showing options like 'Всі', 'Автономна Республіка Крим', 'Вінницька обл.', 'Волинська обл.', and 'Дніпропетровська обл.'. To the right, there are search filters for 'Шукати серед' with checkboxes for 'Всі учасники', 'Емітенти', 'Регістратори', 'Торговці', 'Зберігачі', 'Аудитори', 'Компанії з управління активами', 'Біржі', 'Депозитарії', and 'Анульовані акції'. A 'Пошук' button is at the bottom left. Below the search section is the 'Результати' section, which contains a table of search results.

ЄДРПОУ	Найменування	Область
02470187	Приватне акціонерне товариство "Харківська книжкова фабрика "Глобус" корпоративне підприємство ДАК "Укрвидавполіграфі"	Харківська обл.
20404030	Глобус	Житомирська обл.
20685603	Товариство з обмеженою відповідальністю "Аудиторська фірма "Глобус-аудит"	Автономна Республіка Крим

Figure 42. Primary search on the Smida website (www.smida.gov.ua)

The found statistics of the key market players were downloaded and analyzed.

The screenshot shows the detailed statistics page for the company 'Глобус' (ЄДРПОУ 02470187). The page title is 'Бази даних' and the main heading is '02470187 - Приватне акціонерне товариство "Харківська книжкова фабрика "Глобус" корпоративне підприємство ДАК "Укрвидавполіграфі"'. Below the heading are several tabs for different types of information: 'Інформація про емітента', 'Інформація про випуски', 'Інформація 10 та більше %', 'Архів інформації 10 та більше %', and 'Річна інформація'. The 'Річна інформація' tab is selected, showing a table with columns for 'Дата розміщення', 'Рік', and 'Перегляд'.

Дата розміщення	Рік	Перегляд
04.07.2012 14:40:01	2011	Переглянути
28.04.2011 23:29:55	2010	Переглянути

Figure 43. Statistics of the public company

Step 4. The results of the analysis (steps 1, 2, 3) were compiled into one holistic report of the chosen industry.

Step 5. A final report with all the summaries and economic explanation of the analysis results was written. All the source files were saved.

Practice 6

Knowledge Discovery in Databases (KDD) and data mining

Key goals:

to learn the KDD and data mining algorithms in practice;

to master the technology and algorithms of search of dependences on real data by means of the Deductor Studio or the Longinom Studio.

The theoretical background (in short)

Data mining is a set of computer-assisted techniques designed to automatically mine large volumes of integrated data for new, hidden or unexpected information, or patterns. Data mining is sometimes known as knowledge discovery in databases (KDD) [7].

The capability to deal with voluminous data sets does not mean that data mining requires a huge amount of data as input. In fact, the quality of data to be mined is more important. Aside from being a good representative of the whole population, the data sets should contain the least amount of noise – errors that might affect mining results.

There have been recognized many data mining goals; these goals may be grouped into two categories – verification and discovery. Both of the goals share one thing in common – the final products of the mining process are discovered patterns that may be used to predict the future trends [21].

In recent years, data mining has been studied extensively especially on supporting customer relationship management (CRM) and fraud detection. Moreover, many areas have begun to realize the usefulness of data mining. Those areas include biomedicine, DNA analysis, financial industry and e-commerce. However, there are also some criticisms on data mining shortcomings such as its complexity, the required technical expertise, the lower degree of automation, its lack of user friendliness, the lack of flexibility and presentation limitations.

Data mining methods can be used to perform a variety of tasks, including association discovery, clustering, regression, and classification.

Cluster analysis

Cluster analysis is the best-known descriptive data mining method. Cluster analysis addresses segmentation problems. The objective of this analysis is to separate data with similar characteristics from the dissimilar ones. The difference between clustering and classification is that while clustering does not require pre-identified class labels, classification does. That is why classification is also called supervised learning while clustering is called unsupervised learning.

Sometimes it is more convenient to analyze data in the aggregated form and allow breaking down into details if needed. For data management purpose, cluster analysis is frequently the first required task of the mining process. Then, the most interesting cluster can be focused for further investigation.

There are several ways to perform a cluster analysis. It is therefore important to have a clear understanding of how the analysis will proceed. We can distinguish hierarchical and non-hierarchical methods of group formation. Hierarchical methods allow us to get a succession of groupings (called partitions or clusters) with a number of groups from n to 1, starting from the simplest, where all the observations are separated, to the situation where all the observations belong to a unique group. The non-hierarchical methods allow us to gather the n units directly into a number of previously defined groups.

Choosing the number of groups is of fundamental importance. There is a trade-off between obtaining homogeneous groups, which typically increases the number of groups, and the need for a parsimonious representation, which reduces the number of groups.

Linear regression

In many applications it is interesting to evaluate whether one variable, called the dependent variable or the response, can be caused, explained and therefore predicted as a function of another, called the independent variable, the explanatory variable. We will use Y for the dependent (or response) variable and X for the independent (or explanatory) variable. The simplest statistical model that can describe Y as a function of X is linear regression.

The linear regression model specifies a noisy linear relationship between variables Y and X , and for each paired observation (x_i, y_i) this can be expressed by the so-called regression function, $y_i = a + bx_i + e_i, i = 1, 2, \dots, n$, where a is the intercept of the regression function, b is the slope coefficient of the regression function, also called regression coefficient, and e_i is the random error of the regression function, relative to the i -th observation.

Note that the regression function has two main parts: the regression line and the error term. The regression line can be constructed empirically,

starting with the matrix of available data. The error term describes how well the regression line approximates the observed response variable.

Regression is a simple and powerful predictive tool. To use it in real situations, it is only necessary to calculate the parameters of the regression line, according to formulae, on the basis of the available data. Then a value for Y is predicted simply by substituting a value for X into the equation of the regression line. The predictive ability of the regression line is a function of the goodness of fit of the regression line, which is very seldom perfect.

Classification

Classification is the process of finding models, also known as classifiers, or functions that map records into one of several discrete prescribed classes. It is mostly used for predictive purpose.

Typically, the model construction begins with two types of data sets – training and testing. The training data sets, with prescribed class labels, are fed into the model so that the model is able to find parameters or characters that distinguish one class from the other. This step is called learning process. Then, the testing data sets, without pre-classified labels, are fed into the model. The model will, ideally, automatically assign the precise class labels for those testing items. If the results of testing are unsatisfactory, then more training iterations are required. On the other hand, if the results are satisfactory, the model can be used to predict the classes of target items whose class labels are unknown.

This method is most effective when the underlying reasons of labeling are subtle. The advantage of this method is that the pre-classified labels can be used as the performance measurement of the model.

A **decision tree** is a graphical classification model in the form of a tree whose elements are essentially organized as follows: (a) every internal (non-leaf) node is labeled with the name of one of the predictor attributes; (b) the branches coming out from an internal node are labeled with values of the attribute labeling that node; and (c) every leaf node is labeled with a class.

A decision tree is usually built (or induced) by a top-down, "divide-and-conquer" algorithm. Other methods for building decision trees can be used. For instance, genetic programming can be used to build a decision tree.

While linear and logistic regression methods produce a score and then possibly a classification according to a discriminant rule, tree models begin by producing a classification of observations into groups and then obtain a score for each group.

Tree models are usually divided into regression trees, when the response variable is continuous, and classification trees, when the response variable is quantitative discrete or qualitative (categorical).

Tree models can be defined as a recursive procedure, through which a set of n statistical units are progressively divided into groups, according to a division rule that aims to maximize a homogeneity or purity measure of the response variable in each of the obtained groups. At each step of the procedure, a division rule is specified by the choice of an explanatory variable to split and the choice of a splitting rule for the variable, which establishes how to partition the observations.

The main result of a tree model is a final partition of the observations.

The purpose of dependency analysis, also called **association analysis**, is to search for the most significant relationship across large number of variables or attributes. Sometimes, association is viewed as one type of dependencies where affinities of data items are described (e.g., describing data items or events that frequently occur together or in sequence). This type of methods is very common in marketing research field.

Step-by-step guide

Step 1. Start the data mining software.

It may be the Deductor Studio or the Longinom Studio. They are the versions of one and the same software, provided by the BaseGroup Labs. Both programs have a demo version for educational purposes, so we may use any or both of them in our classes. Their distinctions lie in the user interface (the Deductor Studio has a Russian interface, the Longinom Studio has an English interface) and supported functions (the Longinom Studio lacks a classification algorithm and automatic building of decision trees is not supported). So in order to show all the possible ranges of task execution let's use the Deductor Studio.

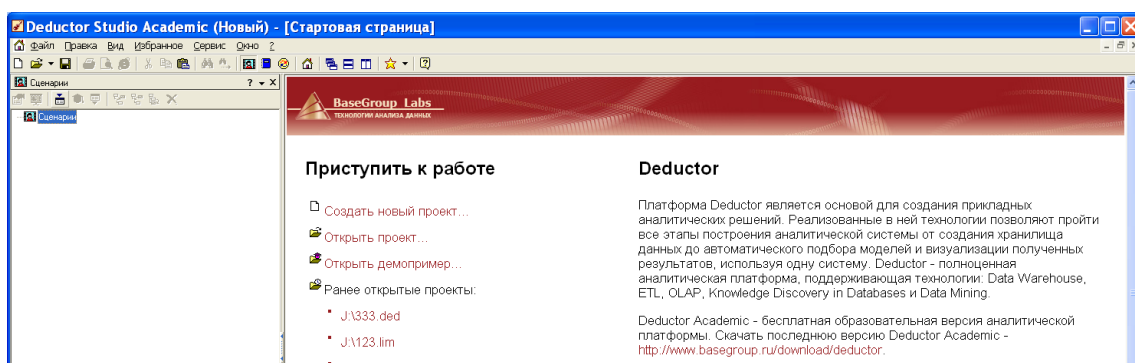


Figure 44. The Deductor Studio starting interface

Step 2. Using "Import master", upload the text file containing data for the analysis.

After the data are checked and displayed correctly, it is possible to pass to the following step.

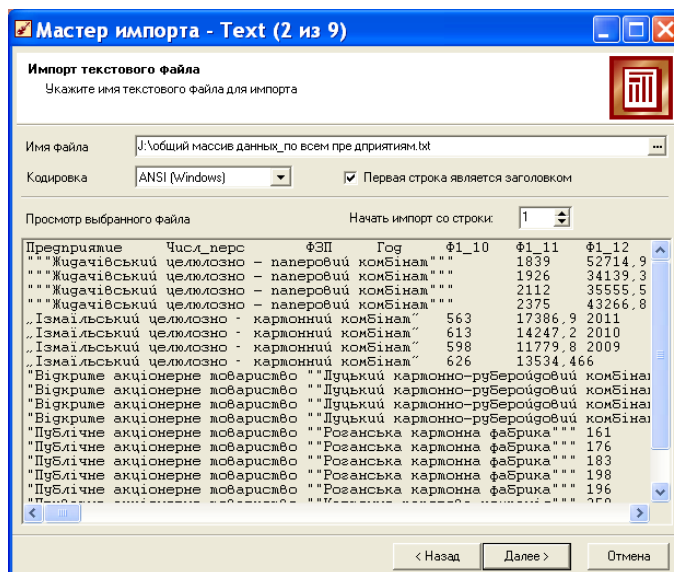


Figure 45. The import wizard (step 2)

Step 3. On the 6th step of importing data from the text file it is necessary to check the correctness of their recognition and purpose. So, all the text data (for example, the names of the enterprises) have to be labeled the "String" type, and all the numerical data are "Real" (generally) or "Integer" (if data in a column can be only integer, for example, as personnel number).

In the "purpose" you should specify that all the columns are "Incoming" as it is supposed that all the columns will be used for calculations.

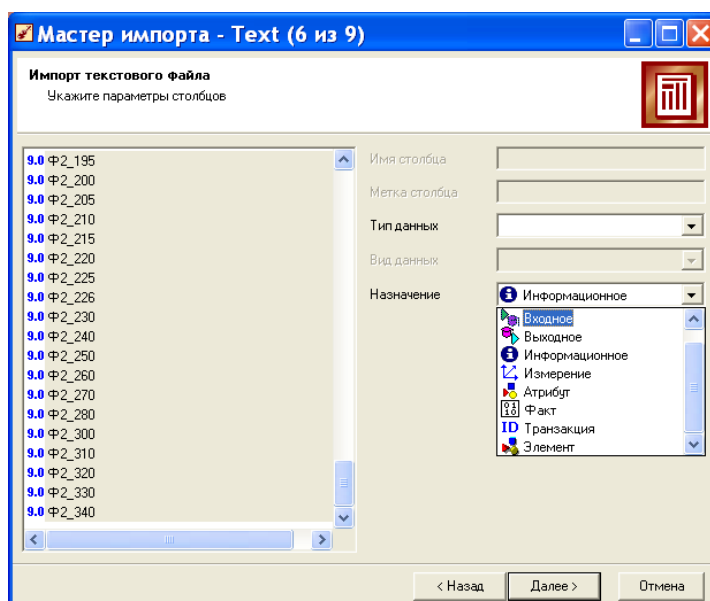


Figure 46. The import wizard (step 6)

Before making calculations it is necessary to **check the data on empty cells**. For this purpose you need to make filtration with the condition of displaying **not empty cells**. Without this operation data will be not available for the analysis.

Step 4. Make the linear regression model.

Let's search for regression dependences on the given data. As the initial file contains data from financial reporting forms, the most "interesting" variable for the analysis is net income from sales (form 2, row 35). The indicators influencing (and defining) profitability of an enterprise are the number of personnel, stakeholders equity (form 1, row 380), material and salary expenses (form 2, row 230), (form 2, row 240), the volume of depreciation charges (form 2, row 260). The list of independent variables can be expanded according to your wishes/assumptions (or proceeding from research objectives).

Let's search for dependence.

4.1. We will click the Processing wizard, choose the "linear regression".

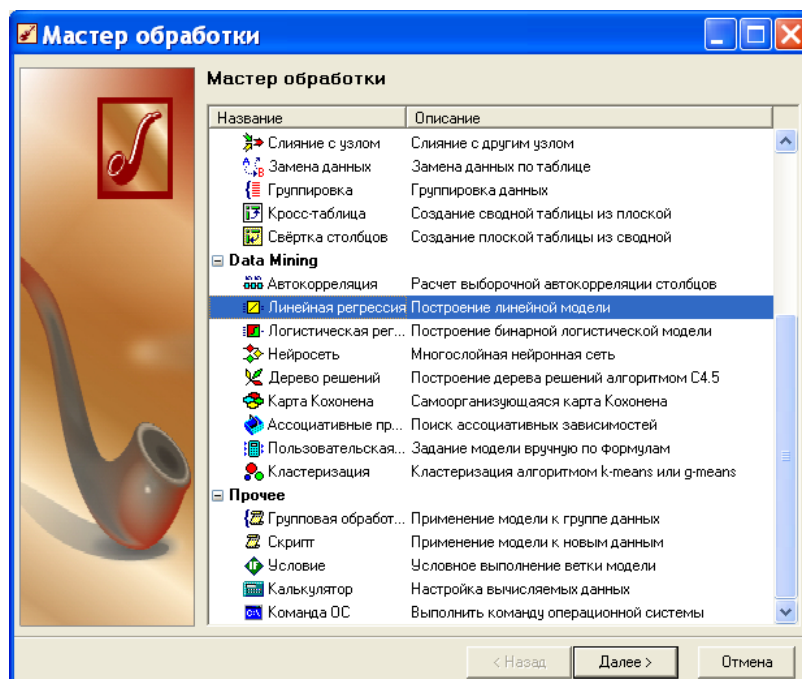


Figure 47. The processing wizard

4.2. When appointing the purposes of initial columns, be sure to specify the variable (as "Outcoming") and parameters (as "Incoming").

Other fields (columns) should be marked as not used.

If necessary it is possible to use restrictions on a dependent variable. This option is used for functions with gaps. In our case it is unreasonable.

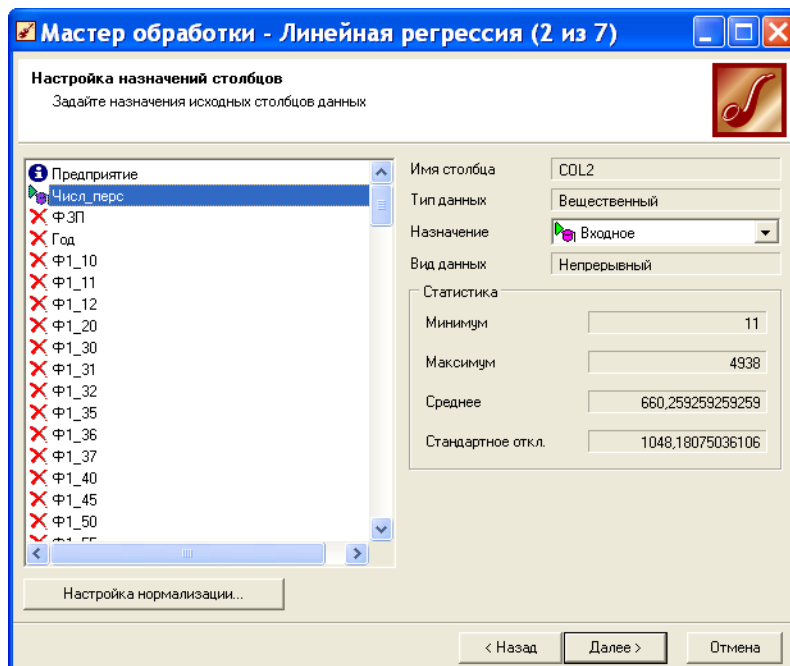


Figure 48. The processing wizard (step 2)

4.3. Let's define the ways of displaying the data. The options "regression coefficients" and "dispersion chart" need to be chosen anyway as they describe the received model and reflect the main characteristics of its quality.

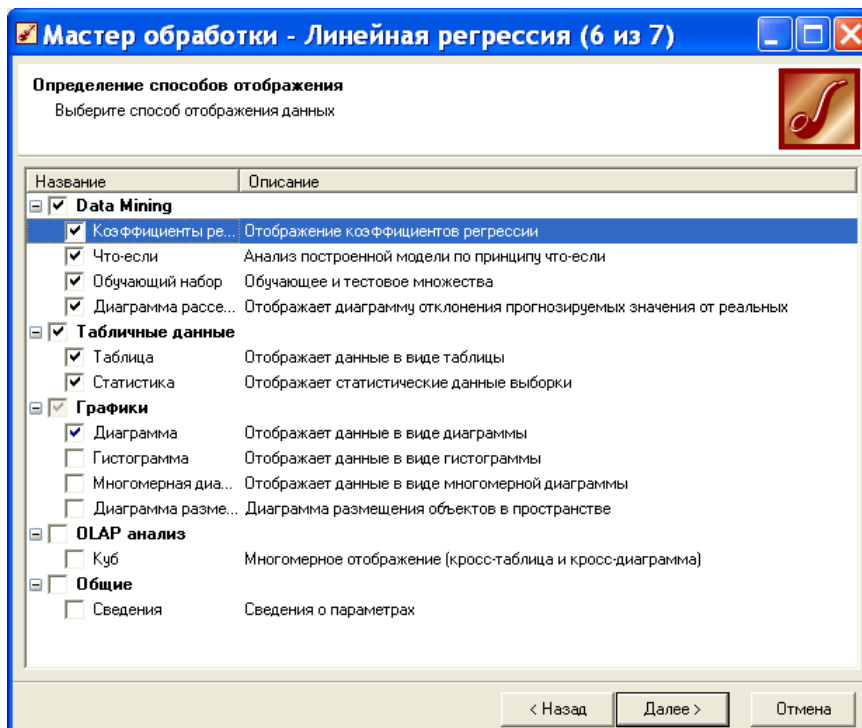


Figure 49. The processing wizard (step 6)

4.4. As a result we have the following model.

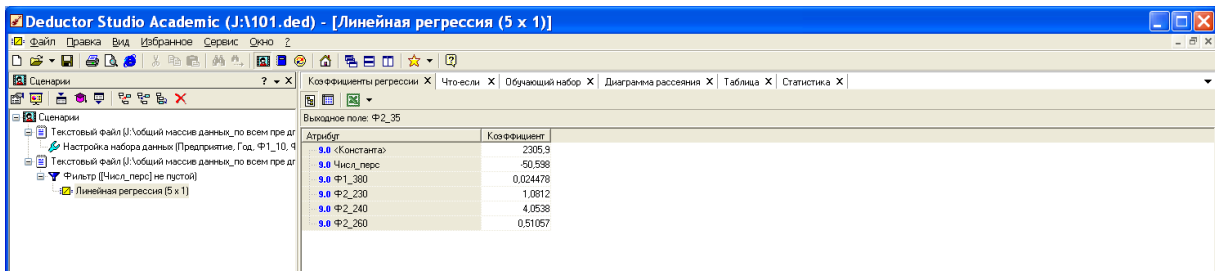


Figure 50. The model

The chart of dispersion reflects a high quality of the model.

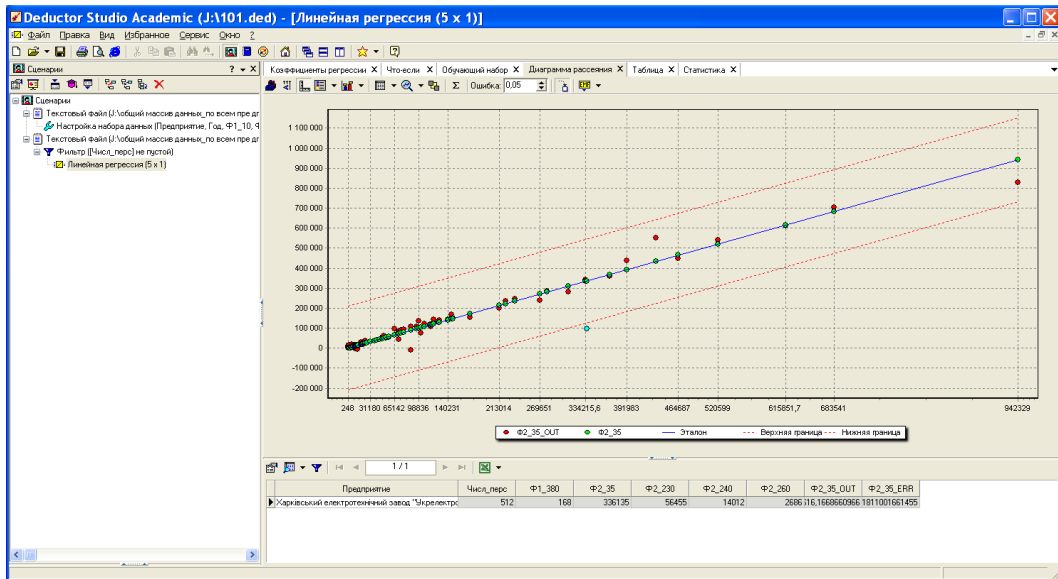


Figure 51. The regression diagram

It is possible to look through the values of any points on the dispersion chart that is especially important in the analysis of deviations in the model.

Choosing the Assessment the Quality of Model button (a sum sign), it is possible to trace an increment of model errors.

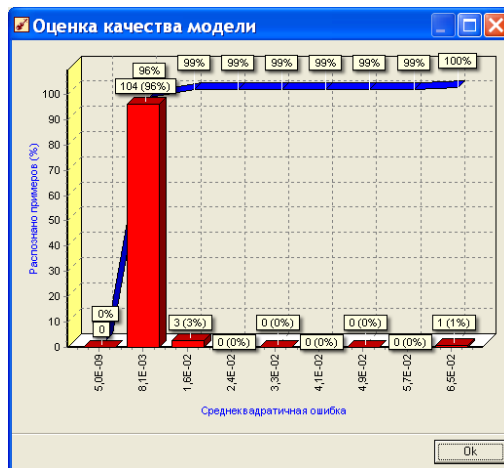


Figure 52. The quality of the model

In the Statistics tab you can see the main information about the incoming data (graphic display can be turned on, using the Statistics Review ("Обзор статистики") button (glasses)).

If necessary the format of data display can be adjusted, using the Fields Settings ("Настройка полей") button. It can be useful when data are displayed in a scientific format, or if it's needed to decrease the accuracy of display to make it clearer (for example, reducing the number of signs after a comma).

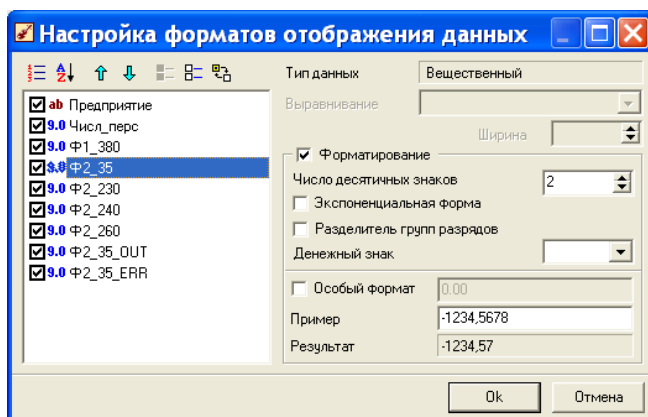


Figure 53. Fields settings

Step 5. Let's make clustering.

On the basis of the data obtained as a result of the previous analysis we can assume that the enterprises can be grouped according to profitability level.

During the setting of fields for clustering it is necessary to specify that an output field is profitability level as we assume that the enterprises vary first of all, by this variable.

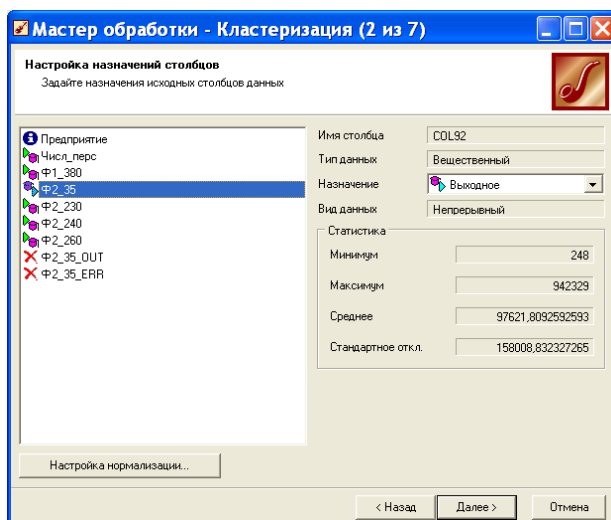


Figure 54. The processing wizard (variables choice)

Clustering can be carried out through specifying the necessary number of clusters, and having provided to be defined by the program automatically.

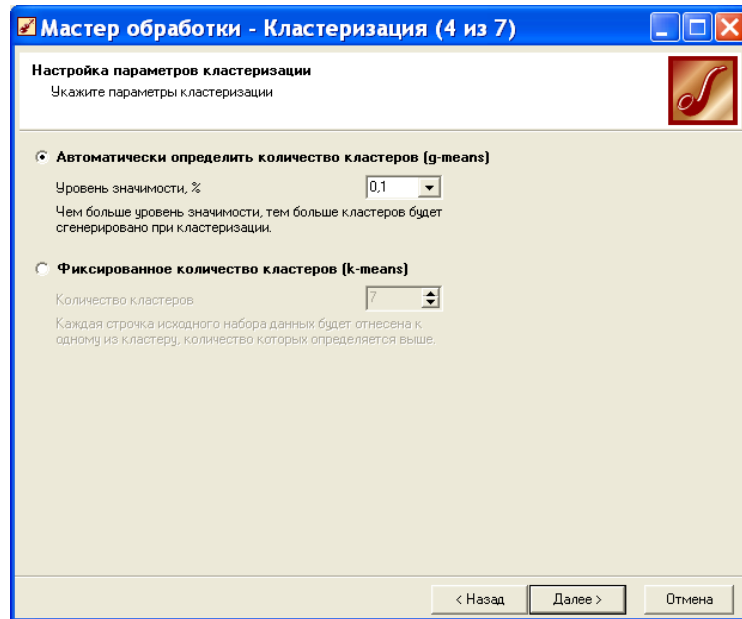


Figure 55. The processing wizard (the choice of the clustering method)

We choose a way of data demonstration.

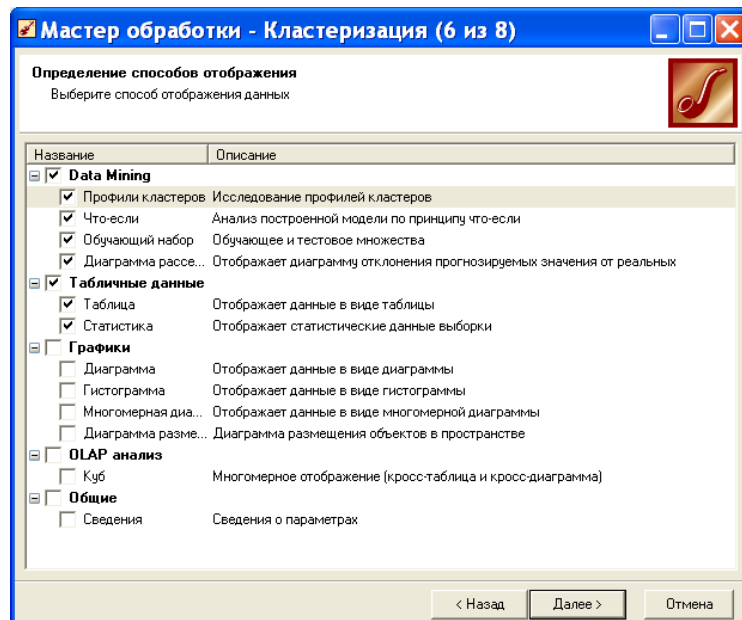


Figure 56. Data presentation settings

As a result of the analysis we've got 3 clusters, where 67, 29 and 12 enterprises are grouped respectively.

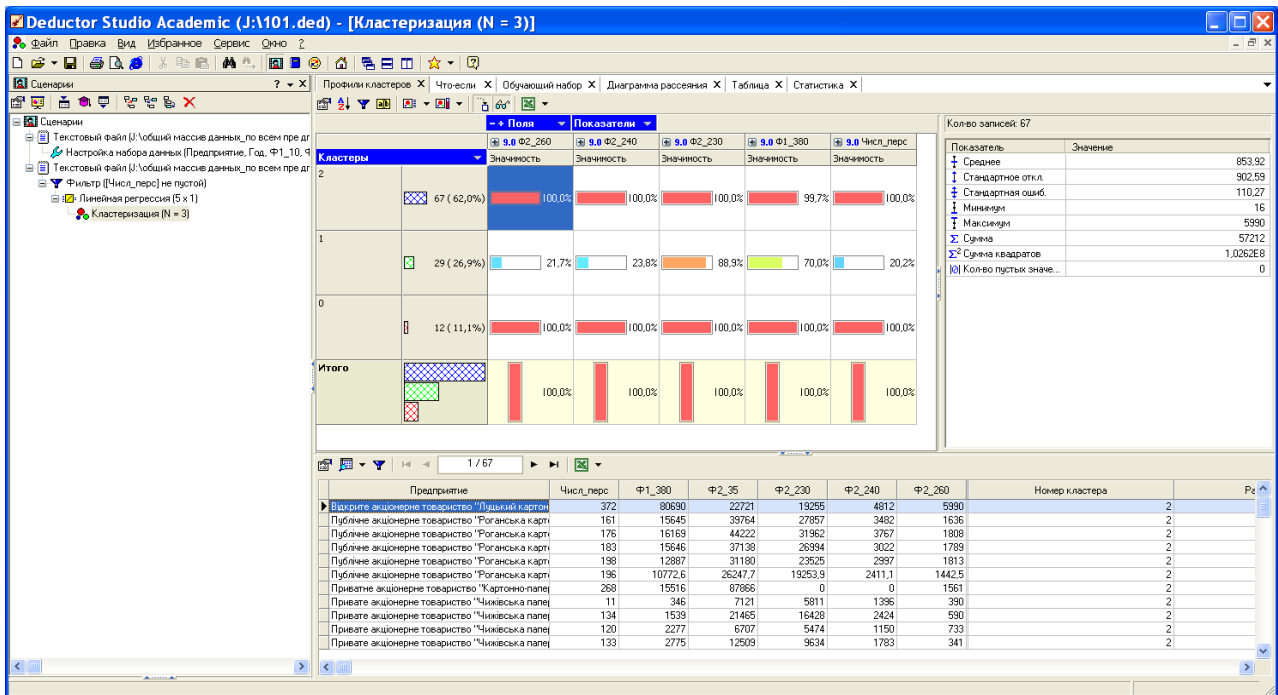


Figure 57. The resulting cluster (the first one)

The first cluster includes the enterprises with low profitability, small and average number of the personnel.

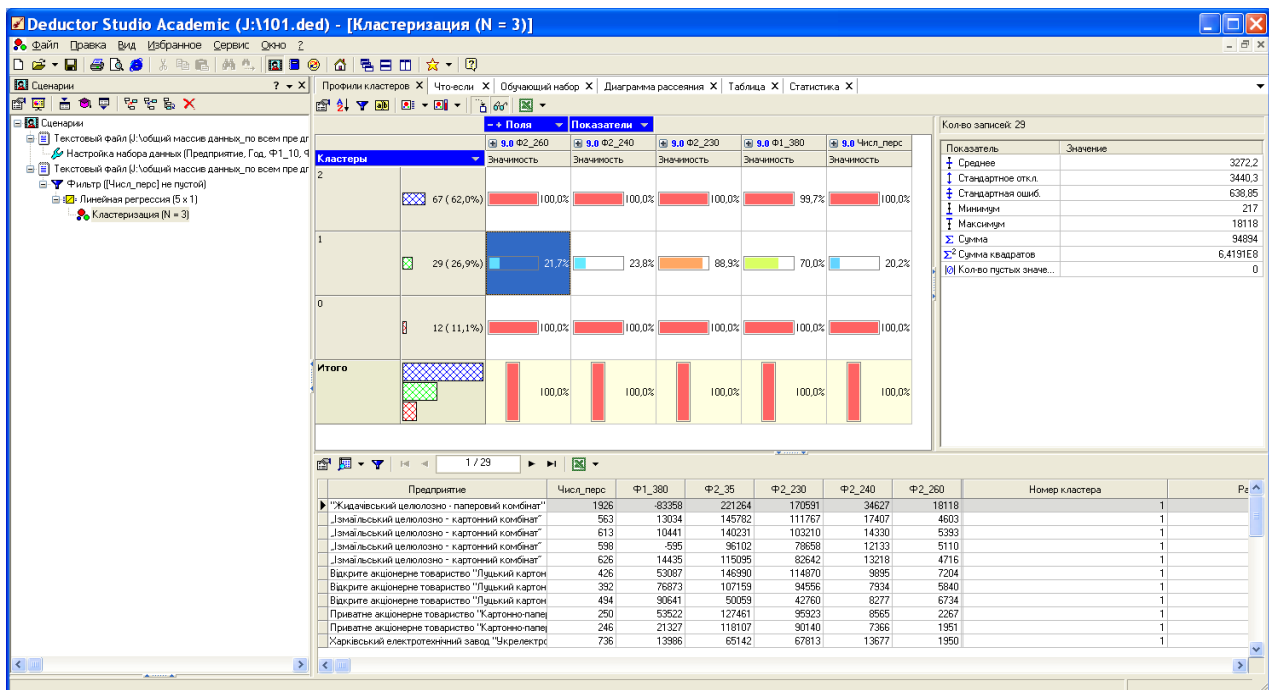


Figure 58. The resulting cluster (the second one)

The second cluster consists of the enterprises with higher profitability, with an average number of the personnel.

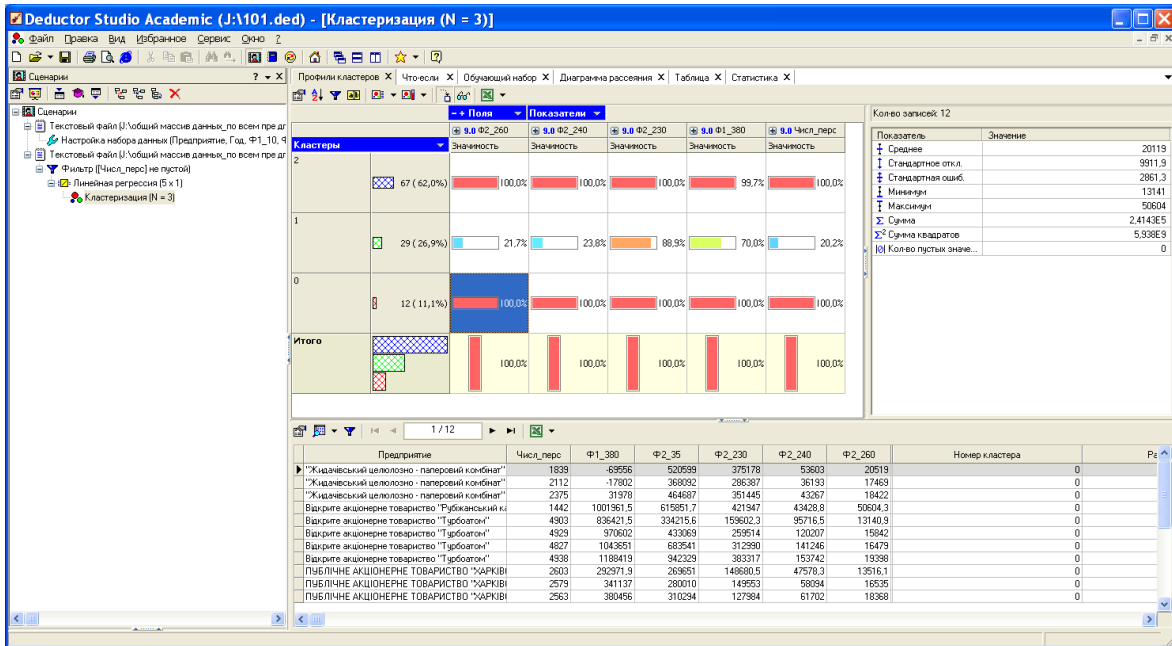


Figure 59. The resulting cluster (the third one)

The third cluster is made of the enterprises with high profitability and a big number of the personnel. The results of the clustering are adequate and correspond to common economic sense.

Step 6. Let's make a decision tree.

In order to identify the most significant signs of division of objects to clusters we'll construct a decision tree. However, it should be mentioned that the decision received as a result can be not unique. ***It is determined with characteristics of a randomly created learning session.*** We choose the algorithm a "decision tree" ("Дерево решений") in the Processing Wizard ("Мастер обработки"). We mark the cluster number variable as an output field.

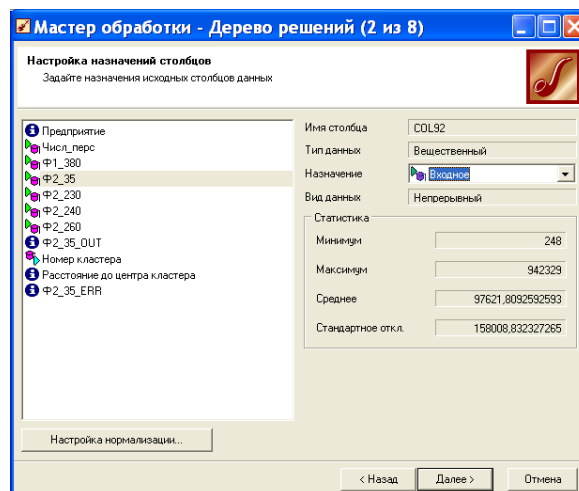


Figure 60. The processing wizard (the choice of variables)

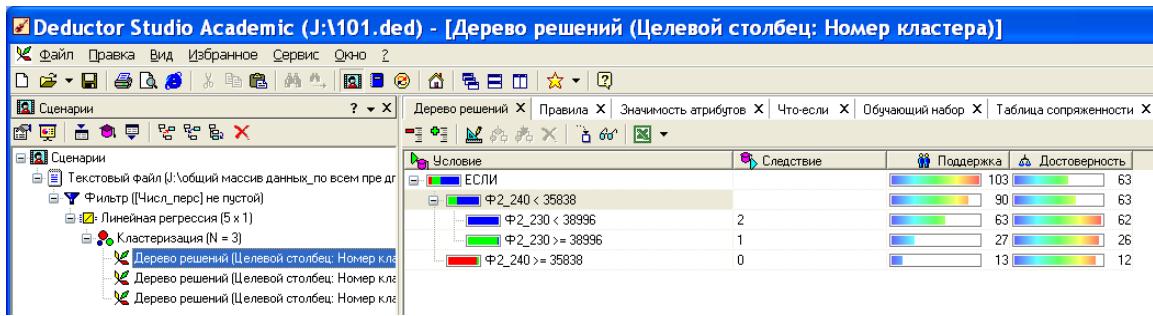


Figure 61. Decision tree dependencies (the hierarchy form)

The decision tree shows the influence of material expenses and personnel cost on the distribution of enterprises in the clusters.

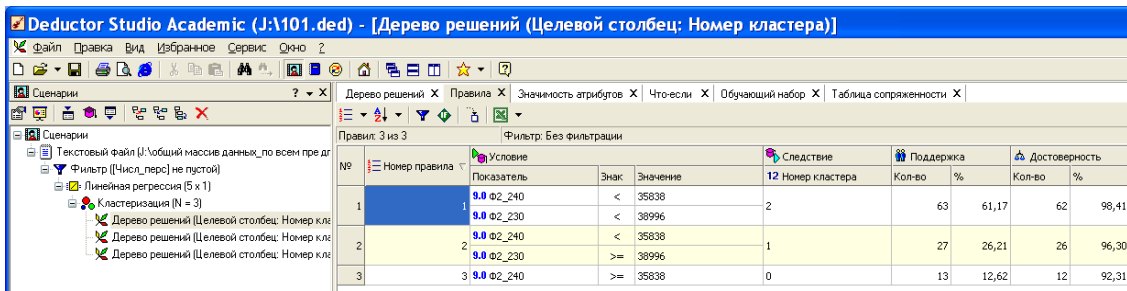


Figure 62. Decision tree dependencies (the equation form)

In order to check the assumption of dependency of enterprises distribution to clusters on the number of the personnel and material expenses, the second decision tree was constructed.

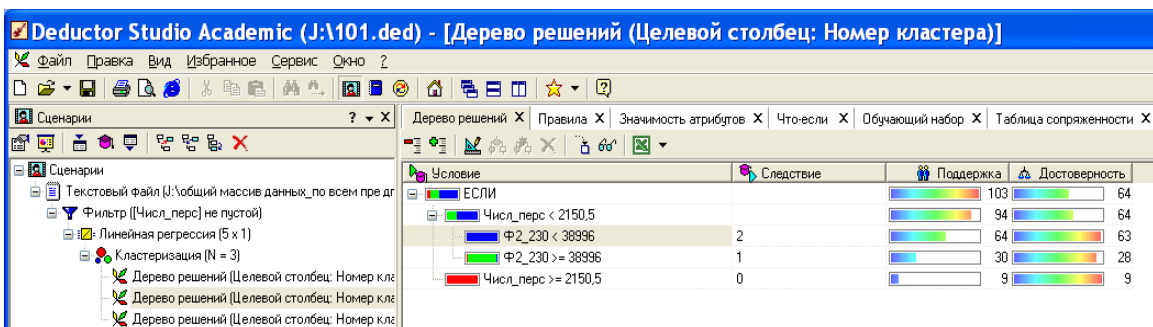


Figure 63. Decision tree dependencies (the hierarchy form)

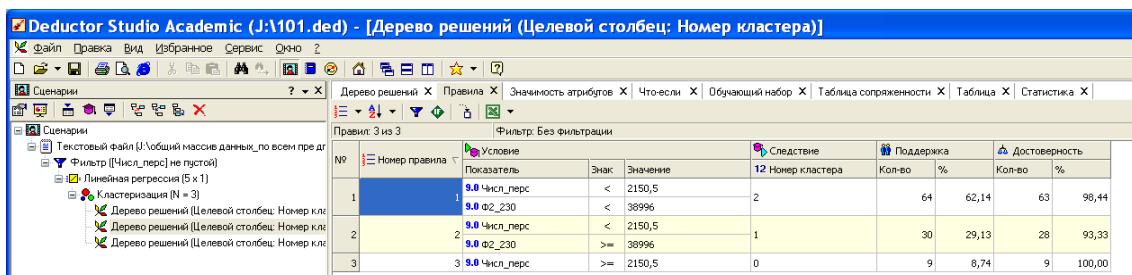


Figure 64. Decision tree dependencies (the equation form)

However, the received decision trees aren't predictive enough and don't suit the research objectives. Therefore, let's construct a decision tree "manually".

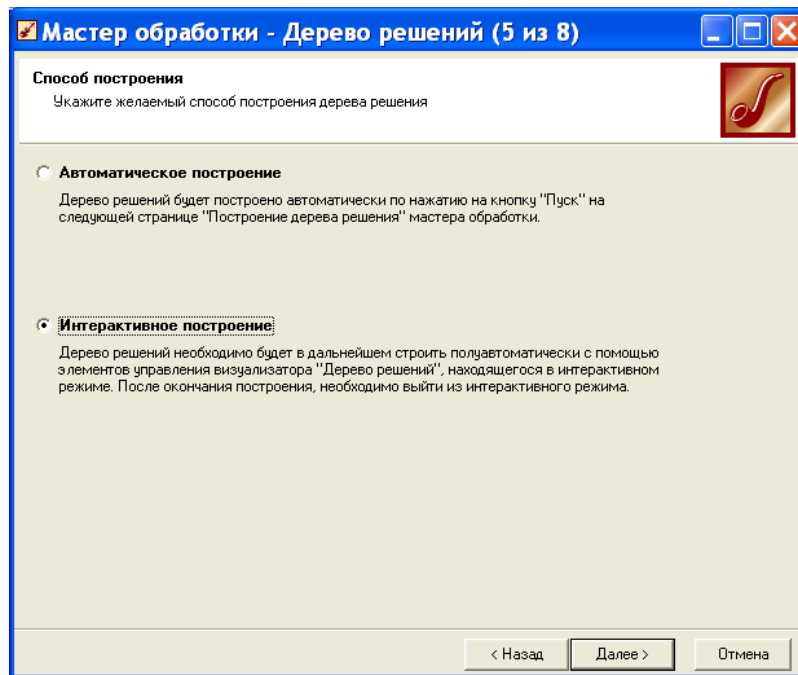


Figure 65. The choice of a decision tree method

Interactive creation of a decision tree requires the usage of the "Break the Current Node into Subsites" ("Разбить текущий узел на подузлы") and "Construct a Tree, Starting with the Current Node" ("Построить дерево, начиная с текущего узла") buttons. The first button allows us to select the fields in which there is the division, the second one automatically builds a tree.

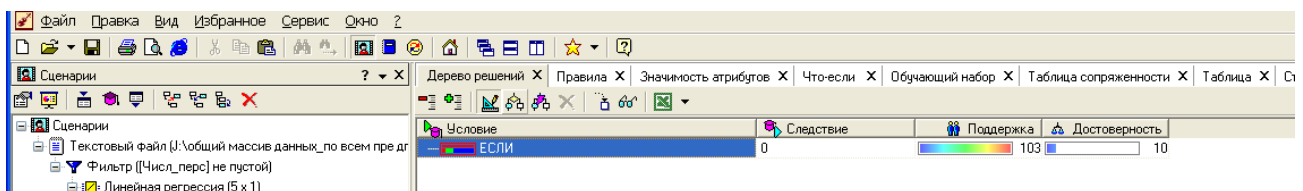


Figure 66. The first step in the "manual" decision tree construction

In order to subdivide the current node into subsites, one uses the manual splitting node ("Break the Current Node into Subsites" (Разбить текущий узел на подузлы) button). Let's try all the possible options of splitting.



Figure 67. Options of splitting on the personnel variable

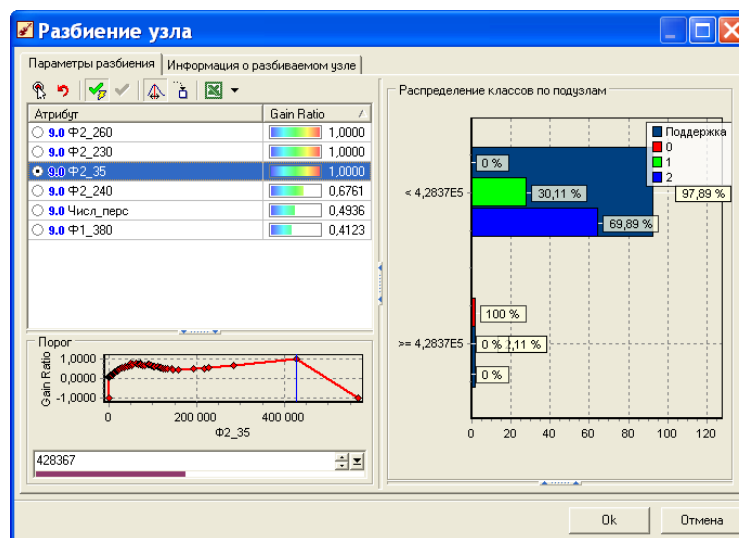


Figure 68. Options of splitting on the revenue variable

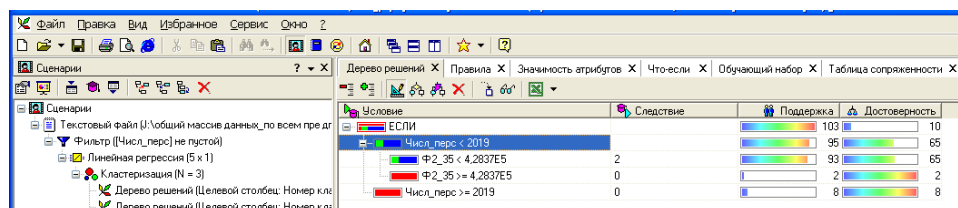


Figure 69. Decision tree dependencies (the hierarchy form)

After two splittings there are only two clusters from the three presented in the Consequences (Следствия) column which means it's necessary to improve the constructed tree as it doesn't reflect all the possible alternatives (clusters) and therefore it is of low quality at this stage.

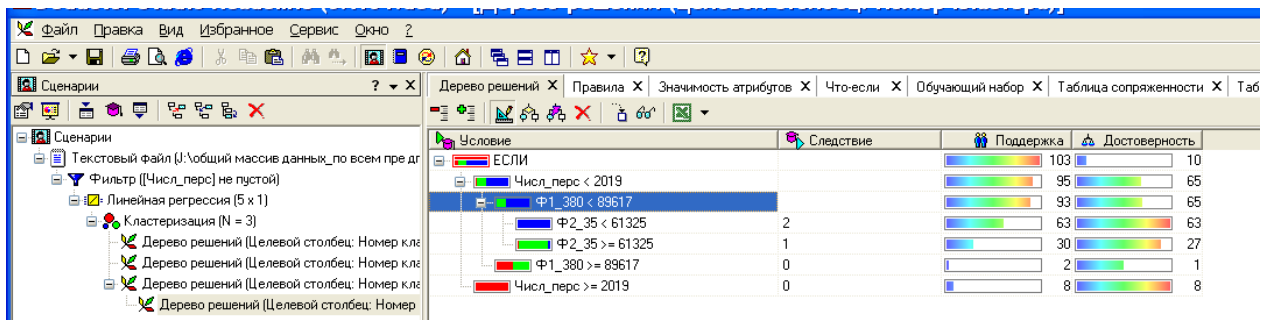


Figure 70. Decision tree dependencies (the hierarchy form)

After a set of improvements the resulting decision tree contains all the three clusters (an additional subdivision of the personnel number < 2019 and the invested capital < 89617; the basis of the division is the enterprise revenue).

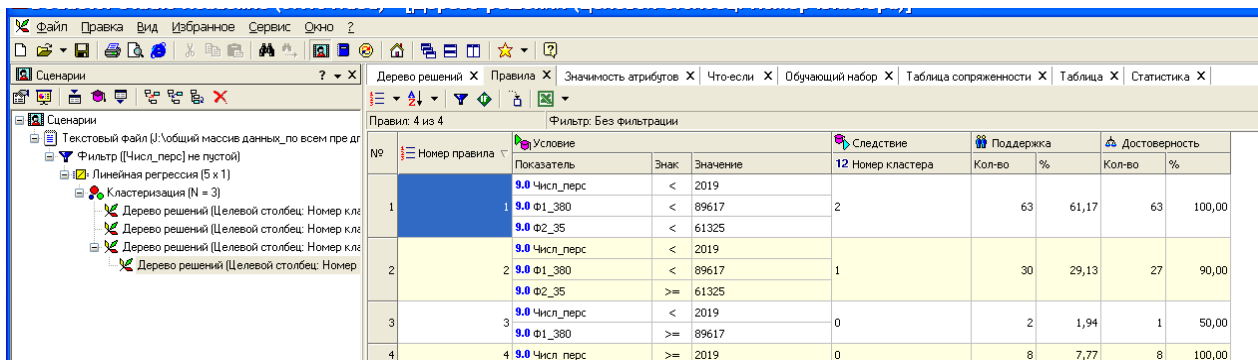


Figure 71. Decision tree dependencies (the equation form)

Finally the received decision tree is full and constructed on the basis of the number of personnel, stockholder's equity (invested in production) and profitability. It corresponds to common (economic) sense and might be meaningfully explained.

Step 7. Save the Deductor Studio file with all the models and the source (data) file. Make a report with all the economic explanations of the results and the summary. Submit your report and the related files.

Practice 7

Knowledge discovery in databases (KDD) and data mining on the individual industry data

Key goals:

to master the technology and algorithms of search of dependences on the real individual data by means of the Deductor Studio.

Step-by-step guide

Step 1. Form a dataset.

During practice 5 all the students searched and aggregated data on a specific industry, so they may be used for this task. All the data should be combined in one meaningful table. As far as all the data are about the industry dynamics, so the columns may contain the industry variables and the rows may have the country names. Another variant is variables placed in rows and the data period (i.e. by year) given in columns.

Step 2. Cleanup the missing data. Save data in a proper format.

The specific feature of the Deductor Studio (as well as the Longinom Studio) is creating a scenario of data processing, not saving the data themselves. So it is highly important to download the data in a proper format. Substitute all the missing values by zeros, delete all non-numerical values from the data part of the table (and substitute them by zeros if necessary).

Due to working with the demo versions of the Deductor Studio (as well as the Longinom Studio), for educational purposes it is obligatory to save the data table in the .txt format.

Step 3. Start the data mining software.

Step 4. Download the data file, check all the values automatically. After checking that the data are displayed correctly, it is possible to pass to the following step. Define the data types.

Step 5. Make a linear regression model on the assumption of the industry variables interdependencies.

Step 6. Make a cluster model on the assumption of the industry variables interdependencies.

Step 6. Make a decision tree on the assumption of the industry variables interdependencies. Don't use all the variables, only a few for the model to be meaningful.

Step 7. Save the Deductor Studio (Longinom Studio) file with all the models and the source (data) file. Make a report with all the economic explanations of the results and the summary. Submit your report and the related files.

References and recommended reading

Main

1. Adam F. Encyclopedia of decision making and decision support technologies / F. Adam, P. Humphreys (editors). – Hershey : Information science reference, 2008. – 976 p.
2. Burstein F. Handbook on decision support systems. In 2 vol. Vol. 1 : Basic themes / F. Burstein, C. W. HoLsapple (editors) ; Berlin : Springer-Verlag Heidelberg, 2008. – 908 p.
3. Burstein F. Handbook on decision support systems. In 2 vol. Vol. 2 : Variations / F. Burstein, C. W. HoLsapple (editors) ; Berlin : Springer-Verlag Heidelberg, 2008. – 856 p.
4. Daellenbach H. Systems and decision making: A management science approach / H. G. Daellenbach. – London : John Wiley & Sons Ltd, 1994. – 545 p.
5. Eom S. Decision Support Systems. International Encyclopedia of Business and Management / S. B. Eom ; edited by Malcolm Warner ; 2nd edition. – London : International Thomson Business Publishing Co, 2001. – 854 p.
6. Laughlin P. Group problem solving / P. R. Laughlin. – Princeton : Princeton University Press, 2011. – 141 p.
7. Maimon O. Data mining and knowledge discovery handbook / O. Maimon, R. Lior (editors) ; 2nd ed. – New York : Springer+Business Media, LLC, 2010. – 1 306 p.
8. Niesbert R. Handbook of statistical analysis and data mining applications / R. Niesbert, J. Elder, G. Miner. – London : Elsevier Inc., 2009. – 824 p.
9. Nutt P. Handbook of decision making / P. Nutt, D. Wilson. – London : John Wiley & Sons Ltd, 2010. – 722 p.
10. Pidd M. Tools for thinking: Modelling in management science / M. Pidd ; 2nd ed. London : John Wiley & Sons Ltd, 2003. – 322 p.
11. Proctor T. Creative problem solving for managers: developing skills for decision-making and innovation / T. Proctor ; 2nd edition. – New York : Routledge, 2005. – 310 p.
12. Pushkar O. Decision support systems : summary of lectures / O. Pushkar, O. Zavgorodnia. – Kh. : Publishing House of KhNUE, 2015. – 172 p.

13. Vercellis C. Business intelligence: Data mining and optimization for decision making / C. Vercellis. – London : John Wiley & Sons Ltd, 2009. – 420 p.

14. Zopounidis C. Handbook of multicriteria analysis / C. Zopounidis, P. M. Pardalos. – Berlin : Springer-Verlag Heidelberg, 2010. – 481 p.

Additional

15. Burstein F. Supporting real time decision-making. The role of context in decision support on the move / F. Burstein, P. Brezillon, A. Zaslavsky. – New York : Springer+Business Media, LLC, 2011. – 403 p.

16. Carutasu G. Decision support system for integrated organizational management / G. Carutasu, C. Botezatu, C. P. Botezatu // Academic journal of manufacturing engineering. – 2008. – Vol. 6, No. 1. – P. 6–11.

17. Christakos G. Integrative problem-solving in a time of decadence / G. Christakos. – New York : Springer+Business Media, LLC, 2011. – 558 p.

18. Christensen H. Cognitive systems / H. I. Christensen, G.-J. M. Kruijff, J. L. Wyatt. – Berlin : Springer-Verlag Heidelberg, 2010. – 494 p.

19. Cizek P. Statistical tools for finance and insurance / P. Cizek, W. K. Hardle, R. Weron ; 2nd edition. – Berlin : Springer-Verlag Heidelberg, 2011. – 425 p.

20. Deemen A. Collective decision making. Views from social choice and game theory / A. Van Deemen, A. Rusinowska. – Berlin : Springer-Verlag Heidelberg, 2010. – 281 p.

21. Han J. Data mining: Concepts and techniques / J. Han, M. Kamber, Jian Pei ; third edition. – New York : Morgan kaufman publishers. 2012. – 703 p.

22. Leleur S. Complex strategic choices. Applying systemic planning for strategic decision making / S. Leleur. – London : Springer-Verlag Limited, 2012. – 168 p.

23. Loshin D. Master data management / D. Loshin. – Burlington : Elsevier Inc., 2009. – 301 p.

24. Managing information technology / C. V. Brown, D. W. DeHayes, J. A. Hoffer et al ; 7th edition. – New Jersey : Prentice Hall, 2012. – 743 p.

25. McFadden D. Rationality for economists? / D. McFadden // Journal of Risk and Uncertainty. – 1996. – No. 8, August. – P. 3–39.

26. McGonagle J. Proactive intelligence. The successful executive's guide to intelligence / J. J. McGonagle, C. M. Carolyn. – London : Springer-Verlag Limited, 2012. – 181 p.

27. Mirkin B. Core concepts in data analysis: summarization, correlation and visualization / B. Mirkin. – New York : Springer, 2011. – 412 p.

28. Ogiela L. Advances in cognitive information systems / L. Ogiela, M. R. Ogiela. – Berlin : Springer-Verlag Heidelberg, 2012. – 119 p.

29. Ritchey T. Wicked problems – social messes. Decision support modelling with morphological analysis / T. Ritchey. – Berlin : Springer-Verlag Heidelberg, 2011. – 112 p.

30. Saaty T. L. Group decision making. The possibility of group welfare functions / T. L. Saaty, K. Peniwati // Information Technology and Decision Making. – 2005. – No. 4(2). – P. 167–176.

31. Taylor J. Decision management systems. A practical guide to using business rules and predictive analytics / J. Taylor. – New York : IBM Press, 2011. – 313 p.

32. Witten I. Data mining. Practical machine learning and techniques / I. Witten ; third edition. – New York : Morgan Kaufman publishers. 2011. – 629 p.

Information resources :

33. Dillon S. Descriptive decision making: comparing theory with practice [Electronic resource] / S. M. Dillon. – [Electronic data]. – Access mode : <http://dssresources.com/papers/ddsscomptheory>. – Title from the screen.

34. Simon H. A. Rational decision-making in business organizations [Electronic resource] / H. A. Simon // Nobel memorial lecture, 8th Dec., 1978. – Access mode : http://www.nobelprize.org/nobel_prizes/economics/laureates/1978. – Title from the screen.

Contents

Introduction.....	3
Practice 1. The decision process stages. The intelligence stage	4
Practice 2. The decision-making process stages. Computer support of the design and choice stages of decision-making	11
Practice 3. Group decision-making. Computer support of group decision-making.....	20
Practice 4. Computer support of group decision-making on individual data.....	27
Practice 5. Data search for industry analysis	28
Practice 6. Knowledge discovery in databases (KDD) and data mining	39
Practice 7. Knowledge discovery in databases (KDD) and data mining on the individual industry data	54
References and recommended reading	56

EDUCATIONAL EDITION

**Guidelines for laboratory works
on the academic discipline
"DECISION SUPPORT SYSTEMS"
for full-time students of training direction
6.030601 "Management"**

Compiled by O. **Zavgorodnia**

Editorial director O. *Pushkar*

Editor Z. *Zobova*

Proof-reader Z. *Zobova*

НАВЧАЛЬНЕ ВИДАННЯ

**Методичні рекомендації
до виконання лабораторних робіт
з навчальної дисципліни**

**"СИСТЕМИ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ"
для студентів напряму підготовки
6.030601 "Менеджмент" денної форми навчання
(англ. мовою)**

Укладач **Завгородня** Ольга Сергіївна

Відповідальний за випуск *Пушкар О. І.*

Редактор *Зобова З. В.*

Коректор *Зобова З. В.*

План 2015 р. Поз. № 80.

Підп. до друку 02.09.2015 р. Формат 60x90 1/16. Папір офсетний. Друк цифровий.
Ум. друк. арк. 3,75. Обл.-вид. арк. 4,69. Тираж 30 пр. Зам. № 142.

Видавець і виготівник – ХНЕУ ім. С. Кузнеця, 61166, м. Харків, просп. Леніна, 9-А
*Свідоцтво про внесення суб'єкта видавничої справи до Державного реєстру
ДК № 4853 від 20.02.2015 р.*