

Government of the Russian Federation
Federal State Autonomous Educational Institution of the Higher Professional
Education

"National Research University

"Higher School of Economics"

Faculty of Business and Management

School of Logistics

Program of the course

«Scientific Workshop»

The section “Methods and tools of data analysis in logistics”

for the educational program «Logistics and Supply Chain Management»

of the training direction 38.03.02 «Management»

Higher education level

BACCALAUREATE

Authors of the program:

Zakhodiakin G.V., senior lecturer, postlogist@gmail.com

Morozova Y.A., candidate of sciences, associate professor, yumorozova@hse.ru

Approved at the department of information systems and technologies in logistics meeting

« 27 » August 2015

Department Head Novikov V.E.

Recommended by the Academic Council of the educational program

« 27 » August 2015, minutes No. 1

Approved « 31 » August 2015

Program Academic Supervisor Dybskaya V.V.

Moscow, 2015

This program cannot be used by other divisions of the university and other institutions of higher education without the permission of the department - the developer of the program.



1 Application area and normative references

This program of academic discipline establishes minimum requirements for knowledge and skills of the student and determines the content and the types of studies and reporting.

The program is designed for lecturers, teaching this discipline, teaching assistants and students of training direction 38.03.02 "Management", studying Bachelor's program "Logistics and Supply Chain Management".

The program is designed in accordance with:

- educational standard of the Federal State Autonomous Educational Institution of Higher Professional Education "National Research University "Higher School of Economics", training direction 38.03.02 "Management";
- working curriculum of the University for the training direction 38.03.02 "Management", Bachelor's program "Logistics and Supply Chain Management", approved in 2015.

2 Objective of the discipline mastering

The section “Methods and tools of data analysis in logistics” is a part of research seminar that aims to study modern methods and tools of visualization, analyzing and forecasting data, and also develop skills of their application in the study and detection of regularities of socio-economic processes for the solution of management problems.

The course focuses on the practical application of data analysis tools. All classes are conducted in a computer lab and include a brief review of the necessary theoretical principles, the method implementation in software and the solution of practical tasks.

3 Student’s competence formed as a result of the discipline mastering

As a result of the discipline mastering the student must:

Know:

- research methods, analysis and detection of patterns in the data;
- methods of forecasting quantitative and qualitative socio-economic indicators;
- methods of classification and clustering of socio-economic objects;
- methods for predicting stochastic phenomena in the economy;
- methods of selection of the forecasting model and assessing its adequacy;

Be able:

- to form the problem of data analysis based on the description of the business problem;
- to determine requirements for the collection of data to address data analysis tasks;
- to choose a method for solving data analysis problems with specificity;
- to apply tools for analysis, visualization and forecasting of economic data;

Have the skills of:

- operational data analysis and data mining;
- forming and solving data mining problems;



- analysis of business information about the company and its environment for decision making;
- analytical reporting using business intelligence tools.

As a result of the discipline mastering the student acquires the following competencies:

Competencies	Code by FSAES/ NRU HSE	Descriptors - the main criteria of the mastering (indicators of results achievement)	Forms and methods of studying, contributing to the formation and development of competence
Able to learn, acquire new knowledge, skills, including in the field other than professional	SK-B1	Studies the material on a given topic	Studying the professional literature and preparation of reports
Able to conduct research, including analysis of the problems, setting goals and objectives, the allocation of the object and the subject of research, choosing research methods, as well as the assessment of its quality	SK-B7	Identifies problems using information about the company Determines the task of data analysis for finding the solution of the problem Selects the methods of analysis and forecasting of economic data, depending on the specifics of a task being solved	Learning teaching materials and answering questions on them The study and discussion of case studies and examples from practice Solving specifically designed tasks Realization of the project
Able to work in a team	SK-B8	Performs the task in the group Participates in group discussion	Group task Group project
Able to analyze and interpret the data of domestic and foreign statistics on the socio-economic processes and phenomena	PK-21	Describes the dynamics of socio-economic indicators Identifies components of a time series (trend, seasonality, cycles)	Learning teaching materials and answering questions on them The study and discussion of case studies and examples from practice Solving specifically designed tasks Realization of the project
Able to select mathematical models of organizational systems, to analyze their adequacy, to adapt models to specific management tasks	PK-22	Selects the model for analysis and forecasting data in accordance with the nature of the data Evaluates the adequacy of the model	Learning teaching materials and answering questions on them The study and discussion of case studies and examples from practice Solving specifically designed tasks Realization of the project
Able to collect, analyze and process data necessary for solving research problems	PK-31	Collects and prepares data for project implementation	Learning teaching materials and answering questions on them Realization of the project
Able to choose tools for information processing according to the	PK-32	Chooses tools depending on the task to be solved Uses tools of analysis and	Learning teaching materials and answering questions on them



Competencies	Code by FSAES/ NRU HSE	Descriptors - the main criteria of the mastering (indicators of results achievement)	Forms and methods of studying, contributing to the formation and development of competence
scientific problem, analyze the results of calculations and substantiate the findings		forecasting data Interprets the results obtained by the tools	Solving specifically designed tasks Preparing reports on executed tasks Realization of the project
Able to analyze and interpret the data of domestic and foreign statistics on the socio- economic processes and phenomena, to identify patterns of changes in socio- economic indicators	PK-34	Describes the dynamics of socio- economic indicators Identifies components of a time series (trend, seasonality, cycles) Identifies the regression relationship between socio- economic indicators	Learning teaching materials and answering questions on them The study and discussion of case studies and examples from practice Solving specifically designed tasks Preparing reports on executed tasks Realization of the project
Able to use modern technical tools and information technologies to solve analytical and research problems	PK-35	Uses tools of analysis and forecasting data Interprets the results obtained using software tools	Learning teaching materials and answering questions on them The study and discussion of case studies and examples from practice Solving specifically designed tasks Realization of the project

4 Place of the discipline in the structure of educational program

Discipline refers to professional cycle and is taught at the 3rd year.

The study of this discipline is based on the following disciplines:

- Information management in logistics
- Probability theory and mathematical statistics
- Economic statistics

Studying the discipline requires following knowledge and competencies:

- basic understanding of data storing and processing;
- knowing the basics of studying the stochastic phenomena in the economy;
- knowing the basic methods of descriptive statistics.

The main provisions of the discipline should be used further in the study of the following subjects:

- Information technologies in logistics and supply chain management
- Inventory management in supply chains
- Information support of logistic activities of trading enterprises
- Sales planning and operations in supply chains
- Audit of the functioning of supply chains
- Cost management in supply chains.



5 Thematic plan of the course

Topic ¹	Total	Class hours	Self-study
Basics of economic forecasting	4	2	2
Data sources, collection and preparation of data for analysis	12	4	8
Exploring data patterns and selecting the forecasting method	8	4	4
Moving average and smoothing methods	6	4	2
Decomposition of time series	6	4	2
Regression analysis	34	18	16
Methods of analysis and forecasting of categorical indicators	28	12	16
Neural networks	6	4	2
Total hours	104	52	52

6 Forms of knowledge control

The course includes visits to academic classes and intensive self-study work. During self-study work students must learn theoretical material to perform the necessary tasks with the use of methodical developments of the Department and recommended literature, to master the technique of working with software products, as well as to carry out the practical exercises and a project dedicated to the solution of analytical problems using studied methods and tools.

Learning management system (LMS) is actively used for organizational and methodological support of the course and contains theoretical materials, practical exercises and examples. Consultations on the implementation of the project are being met through the forum of the course in the LMS.

6.1 Knowledge and skills evaluation criteria

Assessment in all forms of current control are set on a 10-point scale.

7 Contents of discipline

Topic 1. Basics of economic forecasting

The problem of forecasting. Objects of predicting and their properties. Quantitative and qualitative forecasting methods. Types of forecasts. The main stages of the forecasting process. Management of the forecasting process. Computer tools for solving problems of forecasting.

Basic statistical concepts used in forecasting.

Basic books

- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p. Chapters 1-2.
- Chase C. Demand-Driven Forecasting : A Structured Approach To Forecasting. – New Jersey: Wiley, 2009. – 320 p. Chapters 1-2.

¹ The teacher of course may adjust the load on selected topics within the total number of hours for reasons of feasibility and performance of the study group.



Topic 2. Data sources, collection and preparation of data for analysis

Data sources. Conducting surveys. The definition of the studied general population and sample. The selection of survey participants.

Preparing data for analysis. Correction of the data. The heterogeneity of the data sets. The bias of the respondent's characteristics. The problem of outlier answers. Restore missing values.

Basic books

- Norusis M.J. PASW Statistics 18. Guide to Data Analysis, Prentice Hall Inc., 2010. – 653 p. Chapter 3.
- Минько А.А. Прогнозирование в бизнесе с помощью Excel. Просто как дважды два. – М.: Эксмо, 2007. – 208 с. Глава 3.
- Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS [Текст]: учеб. пособие для вузов / А. О. Крыштановский; Гос. ун-т - Высшая школа экономи-ки. — М. : Изд. дом ГУ ВШЭ, 2006. — 281 с. Приложение 2.

Topic 3. Exploring data patterns and selecting the forecasting method

Criteria for assessing data quality. Cross-sectional data and time series. Components of time series: trend, seasonality, cycles. Exploring data using autocorrelation analysis. The model of "white noise". A stationary time series. Selecting the method of forecasting. Assessment of the adequacy of the selected method of forecasting. Tools of exploring data sets.

Basic books

- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p. Chapter 3.

Topic 4. Moving average and smoothing methods

"Naive" forecasting. Simple average. Moving averages. Double moving averages. Exponential smoothing. Forecasting time series with seasonality and trend. Holt's method. Winters' method.

Tools of exponential smoothing and averaging. Examples of implementation of smoothing and moving averages in logistics and supply chain management.

Basic books

- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p. Chapter 4.

Topic 5. Decomposition of time series

Types of trends and trend model fit. The trend forecast. Seasonal coefficients calculating methods. Cyclic and irregular components. The decomposition method, Census II. Deflation.

Tools of time series decomposition.

Basic books

- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p. Chapter 5.

Topic 6. Regression analysis

Simple linear regression. The multiple regression analysis. The multicollinearity. Selection of the regression equation. Regression analysis of time series. Building regression models with autocorrelations. Identification and elimination of autocorrelation. Time series and the problem of heteroscedasticity. Cointegration of time series.

The Box-Jenkins methodology of modeling time series (ARIMA). ARIMA models and its parameters. Bringing the time series to stationarity. The procedure for identification of ARIMA models.



The tools of regression analysis. Examples of regression analysis in logistics and supply chain management.

Basic books

- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p. Chapter 6-9.
- Chase C. Demand-Driven Forecasting: A Structured Approach To Forecasting. – New Jersey: Wiley, 2009. – 320 p. Chapters 3-8.

Topic 7. Methods of analysis and forecasting of categorical indicators

Examining the relationship between categorical variables. Crosstabulation tables.

Factor analysis and its purpose. Analysis of the structure of the data using factor analysis. Data dimension reduction using factor analysis. Interpretation of results of factor analysis.

Cluster analysis. Methods of clustering. Examples of the application of cluster analysis.

Predicting the likelihood of an event. Logistic regression. A likelihood function. Examples of the application of logistic regression model.

Forecasting characteristics of the object using nearest neighbor analysis.

Using decision trees for assessing the risk of occurrence of undesirable events.

Error estimate for the model of forecasting categorical indicators.

Selecting the best model based on ROC curves.

Basic books

- Larose D. Discovering knowledge from data. An introduction to Data Mining – Ch. 10-11.
- Linoff G. Data Mining techniques: for marketing, sales and customer relationship management. – 3ed. – Chapters 13, 14, 18, 19.
- Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных - к знаниям. 2-е изд. - С-Пб. : Питер, 2013. - 704 с. Главы 7-9.
- Наследов А. SPSS 19: профессиональный статистический анализ данных. — СПб.: Питер, 2011. — 400 с. Главы 20-21, 24.

Topic 8. Neural networks

An artificial neural network. Principles of teaching the neural network. The algorithm of back propagation. Application of neural networks in forecasting.

The problem of overfitting models. Assessment of the model error using cross-validation.

Basic books

- Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных - к знаниям. 2-е изд. - С-Пб. : Питер, 2013. - 704 с. - Гл. 8-9.

8 Educational Technology

Educational technology used in the pedagogical process is focused on objectives of learning the discipline that are determined in section 3. It integrates traditional methods of training and active work of students during practical classes with the use of information systems.

Monitoring the quality of learning is accomplished during each training sessions.

Classes are held in computer labs that provide access to specialized software to perform the task by each student.

9 Evaluation tools for monitoring and certification of the student

9.1 Topics of tasks for different types of current knowledge control

◆ Topics of practical exercises

- The calculation of basic descriptive statistics
- Exploring data patterns and selecting forecasting model
- Forecasting using moving averages and exponential smoothing methods
- The fit of the trend model
- Calculation and analysis of seasonal factors
- Simple linear regression model building
- Multiple regression model building and problem of multicollinearity
- Linearization of the dependency of variables using transformations
- Building regression models for time series and identifying the problems of autoregression
- Time series analysis and ARIMA model selection
- Restoring missing data
- Identification of relationships between categorical indicators
- Implementation of factor analysis and interpretation of results
- Implementation of cluster analysis and interpretation of results
- Forecasting the probability of event using logistic regression
- Assessment of risk of undesirable events using decision tree
- Prediction of continuous and categorical indicators using neural network models

◆ Topics of reports based on analysis of publications in scientific and industry magazines

1. The Grammar of Graphics
2. Types of analytical questions & implications for data analysis process
3. Reproducibility in scientific research
4. Correlation, causation and confounding
5. The use of regression in SCM research: factors affecting the level of trust and commitment
6. Combining Judgemental & Statistical Forecasts
7. Demand forecasting for the beverage industry using multi-tiered causal analysis
8. Forecasting competitions: M3 and beyond
9. Tracking forecasting accuracy to improve forecasting process
10. Understanding and preparing data for statistical forecasting
11. Data mining competitions
12. Using artificial neural networks to forecast production cost
13. Using data mining to understand patterns of bike share system usage
14. Using decision trees to predict late arrivals for a transshipment container terminal

◆ Topics of the projects

Students carry out a project dedicated to the solution of analytical problems using studied methods and tools. The project includes following activities. The students analyze information about a selected company and identify the analytical tasks relevant to the solution of its problems. The list of analytical tasks may include forecasting of certain performance, prediction of unknown characteristics of objects according to their known properties, the identification of factors affecting certain indicators and other tasks. Students independently collect the necessary data for their analysis, prepare data for analysis, perform all analyses in accordance with the tasks, form the results in a report. The results of the performed project are reported to a group of students.

9.2 Questions to assess the quality of learning the discipline

- The heterogeneity of the data. Possible causes.
- Methods for recovering missing values.
- Identification of outlier responses and extreme values.
- Time series and its components. Standard decomposition of the time series.
- Types of trends. The fit of the trend model.
- Measures of forecast errors.
- Time series smoothing methods.
- Linear regression model. Key assumptions, the construction procedure and diagnosis.
- Multiple regression model. Key assumptions. The construction procedure and diagnosis. Selecting the factors for inclusion in the model.
- Building a regression model for time-series data. Features associated with the presence of autocorrelations.
- Box-Jenkins methodology (ARIMA) – characteristics and identification of models.
- Using dummy variables for modeling the seasonality and marketing activities.
- Methods of detection of relationship between categorical indicators.
- Factor analysis. The procedure, scope of application, interpretation of the results.
- Cluster analysis. The procedure, scope of application, interpretation of the results.
- Logistic regression model. The scope of application, interpretation of results.
- Nearest neighbor analysis. Examples of usage.
- Forecasting using neural networks.

10 Procedure for the formation of estimates on discipline

The final grade for the discipline consists of the following elements:

- Implementation and presentation of the project
- Activity score (preparation and presentation of oral reports, participation in discussions, performing additional tasks)
- Current knowledge control (solving control tasks, tests)
- A final electronic test (40 min.)

The formula for calculation of final grade for the section of discipline:

$$\text{Final grade} = 0.3 * \text{Score for the project} + 0.2 * \text{Activity score} + 0.1 * \text{Current knowledge control} + 0.4 * \text{Score for the final test}^2$$

The final grade for the research seminar is calculated as the weighted average sum of scores for sections, read each of the Department. Score weights for each section are the proportion of the relevant sections in the total number of completed hours of classroom load.

11 Educational, methodical and information support of discipline

11.1 Basic books

- Chase C. Demand-Driven Forecasting: A Structured Approach To Forecasting. – New Jersey: Wiley, 2009. – 320 p.

² The teacher, leading classes in the study group, may use other weights of the components. In this case the used weight of the components of final grade must be announced to students at the first lesson.



- Hanke J.E., Wichern D.W. Business forecasting. Ninth edition. Pearson Education, New Jersey, 2009. 551 p.
- Norusis M.J. PASW Statistics 18. Guide to Data Analysis, Prentice Hall Inc., 2010. – 653 p.
- Наследов А. SPSS 19: профессиональный статистический анализ данных. — СПб.: Питер, 2011. — 400 с.
- Паклин Н.Б., Орешков В.И. Бизнес-аналитика: от данных - к знаниям. 2-е изд. - С-Пб. : Питер, 2013. - 704 с.

11.2 Additional books

- Larose D. Data Mining Methods and Models. – Wiley, 2006.
- Larose D. Discovering Knowledge in Data: An Introduction to Data Mining. – Wiley, 2005.
- Westerman P. Data Warehousing: Using the Wal-Mart Model. - Morgan Kaufmann, 2000
- Yau N. Visualize this. The FlowingData Guide to Design, Visualization and Statistics. – Wiley, 2011. – 384 p.
- Дебок Г., Кохонен Т. Анализ финансовых данных, Москва, Издательский дом «Альпина», 2001 г.
- Дейвенпорт Т., Харрис Дж. Аналитика как конкурентное преимущество. Новая наука побеждать. - М.: BestBusiness Books, 2010
- Чубукова, И. А. Data Mining. М. БИНОМ. Лаборатория знаний, 2008. - 382 с.

11.3 Remote support of the discipline

LMS is used to ensure an interactive and continuous learning process. The system is used for posting of course materials, execution of projects, testing knowledge and consultation in project implementation through the forum of course

12 Inventory and logistics support of discipline

Classes are held in classrooms equipped with multimedia tools for presentations of reports and project works and computers with installed specialized software: SPSS, R, Rapidminer, and Tableau.