МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ ІМЕНІ СЕМЕНА КУЗНЕЦЯ

"ЗАТВЕРДЖУЮ"
Заступник керівника
(проректор з науково-педагогічної роботи)
_____
М. В. Афанасьєв

СТАТИСТИЧНЕ МИСЛЕННЯ ДЛЯ НАУКИ ПРО ДАНІ

робоча програма навчальної дисципліни

Галузь знань	**12 «Інформаційні технології»**
Спеціальність	**122 «Комп'ютерні науки»**
Освітній рівень	**другий (магістерський)**
Освітня програма	**«Бізнес-аналітика та інформаційні системи в підприємництві»**

Вид дисципліни	**вибіркова**
Мова викладання, навчання та оцінювання	**англійська**

Завідувач кафедри економічної теорії,
статистика та прогнозування О.В. Раєвнєва

Харків
ХНЕУ ім. С. Кузнеця
2018

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

STATISTICAL THINKING FOR SCIENCE ABOUT DATA

syllabus of the discipline

Area of Education	**12 “Information Technology”**
Speciality	**122 “Computer Science”**
Educational level	**second (master's)**
Educational program	**Business Analytics and Information Systems in Entrepreneurship**

Type of discipline	**selective**
Language of teaching, training and evaluation	**foreign (English)**

Kharkiv
S. KUZNETS **KhNUE**
2018

The syllabus has been approved by the Department of Economic Theory, Statistics and Forecasting

Protocol № 4 on 11.10.2018

Compiled by: Rayevnyeva O., doctor of economic sciences, professor of the Department of Economic Theory, Statistics and Forecasting;

Stryzhychenko K., doctor of economic sciences, professor of the Department of Economic Theory, Statistics and Forecasting;

Derykhovska V., candidate of economic sciences, associate professor of the Department of Economic Theory, Statistics and Forecasting

Letter of renewal and re-approval of the syllabus of the academic discipline

Academic year	**Date of the session of the department - developer of syllabus**	**Protocol number**	**The signature of the head of the department**

1. Introduction

Annotation of the discipline:

The rapid development and wide application of the newest packages of applied programs and computer technology tools necessitate the formation of a specialist in business intelligence and information systems of new competencies aimed at acquiring knowledge and skills in using econometric and mathematical modeling for the analysis of complex, mass socio-economic phenomena and processes in various spheres of activity.

Moreover, the sharp increase in information in every area of business and public life, from the environment to diverse market research, is the main driver for creating an unprecedented global demand for information technology specialists who have deep knowledge of business intelligence and work with Big Date At present, the demand for specialists who combine competence in the application of intelligent information and computer systems processing systems, the use of modern software products, IT technologies and technological tools in professional, in particular, entrepreneurial activity with the competencies of business analyst for substantiation and adoption management and business decisions are the main trend in demand in the national and international labor markets. This discipline is a response to the contemporary needs of the community and provides students with an in-depth understanding of the business context of any socioeconomic processes and will enable them to solve the problems associated with analytical work in the IT industry.

Economic systems studied by modern science, with great difficulty, are subjected to the study by ordinary (verbal) theoretical methods. A direct experiment on them is impossible. The price of errors and miscalculations is large, so mathematical modeling is an inevitable component of scientific and technological progress.

Statistical thinking for science about data is one of the basic disciplines of the master's program "Business Analysis and Information Systems in Entrepreneurship".

"Statistical thinking for science about data" is a discipline that studies the tools of economic and mathematical modeling in the study of complex socio-economic systems.

The object of the discipline is complex socio-economic systems.

The subject of the discipline is theoretical and practical questions concerning the development of forecasts and the construction of models in a market economy based on the use of modern economic and mathematical methods and models.

Purpose of the discipline: is the extension and deepening of theoretical knowledge and acquisition of professional competencies of forecasting of socio-economic processes and modeling of complex systems using statistical methods and models.

Course	**1M**	
Semester	**2**	
Number of ECTS credits	**5**	
Audit lessons	**lectures**	**20**
	laboratory	**20**
Independent work		**110**
Form of final control	**examination**	

Structural-logical scheme of studying the discipline:

Previous disciplines	The following disciplines
Mathematics for economists	Economic analysis
Macroeconomics	Business analytics
Statistics	Trainings
Microeconomics	Writing master's theses
Probability Theory and Mathematical Statistics	Disciplines of the second semester of the curriculum of the master's program "Business Analysis and Information Systems in Entrepreneurship".
Computer science	

2. Competence and outcomes of studying in a discipline:

Competence	Results of studying
The ability to acquire theoretical knowledge of statistical modeling and forecasting and receive skills of formation information space research	- define object modeling, choose the type of economic-mathematical models necessary for solving applied problems and form a request to her building; - conduct primary research information space analysis, check the law of distribution of source data series
The ability for modeling relationships between economic processes and phenomena	- conduct the evaluation of the parameters of the linear regression model, determine its adequacy and using models to predict the change in factor signs; - conduct evaluate the regression model for multicollinearity and eliminate it as needed
The ability to model and predict time series in the study of the dynamics of the development of socio-economic systems	- to conduct a decomposition of a number of dynamics and to model its trend component in order to forecast the development of systems (enterprises, regions, etc.); - to detect and simulate cyclic fluctuations in economic processes with the help of a Fourier series expansion; - the ability to construct autocorrelation models and to check the presence of autocorrelation of the model's remnants and its elimination; - to apply adaptive forecasting methods in the study of perspective and retrospective trends in the development of socio-economic systems (SES) (enterprises, organizations, regions, etc.); - use integrated models of autoregression in the study of the economic processes of stochastic nature
The ability to identify and simulate the behavior of homogeneous complex socio-economic systems	- use the method of the main components for reducing the factorial space of research SES; - to group SESs by methods of natural and artificial clusterization; - substantiate the choice of measure of the distance between objects; - to develop a discriminatory function and to check the quality of the clustering of SES

3. Program of the academic discipline

Content module 1. Methodological bases of statistical modeling and forecasting

Topic 1. Categorical basis of statistical modeling and forecasting

1.1. The concept of socio-economic systems, their structure as an object of modeling

Category "system" and its features. Socio-economic systems and their features. Types of structures of socio-economic systems.

1.2. Economics as a subsystem of nature and society

The role of modeling in modern economic theory. Basic theoretical positions of modeling. Stages of the simulation process. Interconnection between economics and mathematics. Features of economic systems that complicate their modeling. Advantages of using mathematical modeling in economics.

1.3. Classification and stages of construction of economic and mathematical models.

The main stages of building models. Classification of economic and mathematical models on various grounds.

1.4. Requirements and principles for the construction of models.

Basis requirements for the construction of economic and mathematical models. The three main levels of the modeling hierarchy. Principles to be met by a mathematical model for use in the economy.

1.5. Prediction as a method of predicting socio-economic processes

Forecasting as a method of predicting socio-economic processes (approaches to solving predictive problems, ways of forecasting development, forms of prediction, prediction functions). Prediction as a management function. Classification of forecasts. Typology of forecasting methods. Stages of statistical forecasting process.

1.6. Contents of the basic categories of forecasting

Basic categories of socio-economic forecasting and their content. Comparison of the features of the basic methods of forecasting. Classification of types of socio-economic objects of forecasting and forms of presentation of results of forecasting. Characteristics of forecast quality, methods of its estimation and classification of forecast errors.

Laboratory work 1 “Introduction to the Statistica 10.0 package. Study of the statistical characteristics of the variation series”. Tasks: fixing theoretical and practical material for the study of statistical characteristics of the variation series in the package Statistica 10.0. Obtaining the skills of the study variation series in the module Descriptive statistics.

References. Main: 3; 6; 9. Additional: 16; 17.

Topic 2. The regression models as a means of researching economic processes

2.1. The concept of multiple regression as a class of econometric models of models and approaches to their construction

Concept of model and econometric model. Types of representations of econometric models. Causes of stochasticity of econometric models. A dual and multiple regression model. Prerequisites for constructing an econometric model. Stages of model construction, their content. Approaches to constructing a multiple regression model.

2.2. Using MNC to calculate model parameters.

The least squares method (MNC) for one factor and multi-factor models. Properties of estimations of model parameters at MNC.

2.3. Checking the quality of built models.

Checking the statistical significance of the model parameters. Student Criterion. Checking the adequacy of the models. Determination coefficient, multiple and pair correlations. Fisher's Criterion. The system of criteria for model quality. Trust intervals when forecasting.

2.4. Multicollinearity, methods for checking and exclusion

Multi-linearity, its kinds. Causes of Multicollinearity. Criteria for verifying the presence of multicollinearity. Methods of eliminating multi-liner.

Laboratory work 2 “Construction of a single-factor and multiple regression model”.

Tasks: fixing theoretical and practical material, acquiring skills for constructing and analyzing simple econometric models in the Statistica 10.0 package. Obtaining skills to analyze the significance of model parameters and their quality. Mastering the basics of analysis of model errors, as well as the features of constructing interval and point prediction in the Multiple Regression module. Acquiring the skills of testing the model for multicollinearity and eliminating its consequences. Domain testing methods for multicollinearity in the model in the MS Excel package. Learn how to build step-by-step inclusion and exclusion models in the Statistica 10.0 package.

References. Main: 2; 6; 7. Additional: 16 – 17; 20.

Topic 3. Modeling and forecasting development of trends

3.1. Concept of time series, its components and classification of time series analysis models.

The concept of the time series, its difference from a number of dynamics. Types of time series. Components of time series: trend, seasonal, cyclic and random components. Methods of analysis of time series: trend models, decomposition models, smoothing models, auto regression and lag models.

3.2. Methods for determining the presence of a trend in dispersion and average: Foster-Stewart method, Fisher method, method of averages.

Determination of the presence of a trend in the dispersion of time series values using the Fisher F-criterion method. The essence of the Foster-Stewart model. Algorithm of the method of averages. Combinations of the presence of the trend in the variance and the average.

3.3. Models of time series decomposition.

Algorithm for constructing a decomposition model, the content of its steps. Aditive, multiplicative decomposition models.

3.4. Autocorrelation, methods of its determination, stationary time series.

The concept of autocorrelation, the causes of its occurrence. Consequences of autocorrelation. Criteria for determining the presence of autocorrelation: Darbin-Watson's criterion, von Neumann's criterion, non-cyclic autocorrelation criterion, cyclic autocorrelation criterion.

Ways to eliminate or reduce autocorrelation: the Frish-Woo method, the method of finite difference, the method of deviations of empirical values from trend alignments, the method of the Cochrane-Orkata method, the Hildret-Lou method.

Stationary time series. Dicci-Fuller's test for time series stationary. Types of test: a simple DF test, an advanced ADF test.

3.5. Criteria for checking the quality of built models.

Criteria for checking the quality of the built models of forecasting of time series: variance, average relative error of approximation, coefficient of ascent, determination coefficient, average square error, average absolute error, average absolute error in percent, coefficient of Tail discrepancy.

Laboratory work 3 “Construction of a time series decomposition model”. Tasks: fixing theoretical and practical material, acquiring skills in the decomposition of the time

series in the package Statistica 10.0. Study of methods for checking the presence of a trend component in a series. Mastering the skills of choosing a trend type using the MS Excel package.

References. Main: 1; 4; 8. Additional: 11; 16 – 17.

Topic 4. Models of adaptive forecasting and integrated model of autoregression

4.1. The concept of anti-aliasing and its types.

The concept of anti-aliasing. General smoothing methods: simple smoothing methods, exponential smoothing, adaptive smoothing. The algorithm for finding the optimal smoothing parameter.

4.2. Adaptive smoothing for Brown, Holt, Winters.

Concept of adaptive smoothing models. Smoothing by Brown, Holt and Winters.

4.3. Integrated model of autoregression.

Concept of the model of autoregression. Mixed model of autoregression. Model of auto regression and integrated moving average.

4.4. Vector autoregressive model

Vector autoregressive model, classification of vector models. Granger's Cause Pulse analysis.

Laboratory work 4 “Estimation of the presence of autocorrelation of the remnants of the model. Eliminating autocorrelation” Task: acquiring the skills of predicting time series using exponential smoothing in the Statistica package 10.0.

References. Main: 4; 8. Additional: 12; 14 – 17.

Content module 2. Modeling and forecasting of multidimensional processes

Topic 5. Factor analysis of data

5.1. Basic concepts of factor analysis

The concept of factor analysis. The main tasks of factor analysis. Requirements for the implementation of factor analysis. Concept factor, vector factor load, reduced matrix.

The notion of community, the problem of community. Evolution of problems of factor analysis. General and special classification of factors. The notion of a general factor. The notion of complexity. Factor analysis algorithm. Classification of types of factor analysis.

5.2. Methods of factor analysis.

Classification of methods of factor analysis. Simplified methods approximating methods with enhanced approximating properties.

5.3. The method of the main components.

Components of factor analysis. The method of the main components, its stages. Methods of constructing a reduced correlation matrix. Matrix factor loader. Criteria for evaluating the required number of factors - Kaiser's criterion, rocky oscillation, correlation matrix.

Laboratory work 5 “Building a model of factor analysis” Task: acquiring data processing skills using factor analysis methods in the Statistica 10.0 package.

References. Main: 5; 10. Additional: 12; 15; 17; 20.

Topic 6. Cluster analysis as a means of forming homogeneous data groups

6.1. The essence of cluster analysis.

The term cluster analysis and its occurrence. Cluster as the basic concept of cluster analysis. Mathematical characteristics and types of clusters. Areas of use of cluster

analysis. Classification and its types. Comparison of clustering and classification. Stages of cluster analysis. Advantages and limitations of cluster analysis.

6.2. Standardization and standardization.

The notion of standardization and standardization. Standardization methods.

6.3. The concept of distance.

Types of measure of identity of objects: correlation coefficient, associativity coefficient, probability coefficient of identity, distance measure. The concept of distance, the classification of distance measures.

Types of quality scales (measure of distances), nominal scales (measure of identity), loose scales, their characteristics.

6.4. Methods of cluster analysis

The concept of a cluster method. Methods of cluster analysis - hierarchical and non-hierarchical methods of clusterization. Hierarchical sgm. And divisimic methods. Rules for merging into a cluster. Medium method, PAM method. Checking the quality of clustering.

Laboratory work 6 “Using cluster analysis for the study of economic processes” Task: getting the skills of using cluster analysis in the Statistica 10.0 package.

References. Main: 5; 10. Additional: 12; 16 – 17; 20.

Topic 7. Data Recognition and Discriminatory Analysis

7.1. Basic concepts of discriminatory analysis.

The concept of discriminatory analysis, types of discriminatory analysis. The notion of a discriminant variable, approaches to the introduction of discriminant variables. Ways of a step-by-step discriminant analysis. The qualities of discriminant variables. The task of discriminant analysis. Discriminatory analysis methods.

7.2. Discriminatory functions.

The notion of discriminating function. Concept of linear and canonical discriminatory function. Principles of finding the coefficients of canonical discriminant functions.

7.3. Criteria for assessing the quality of the classification.

Criteria for evaluating the quality of discriminant functions: Mahalanobis distance, Wilks λ-statistics, criterion and F-statistics.

7.4. Use of discriminatory analysis in the economy

Examples of using discriminant analysis for object recognition. The relationship between discriminant variables and discriminatory functions.

Laboratory 7 ‘’Solving the problem of classification by the method of discriminant analysis” Task: getting the skills of using discriminant analysis in the Statistica 10.0 package.

References. Main: 5; 9; 10. Additional: 13 – 15; 17; 18; 20.

4. The order of assessment of the results of training

The assessment system of students’ competences is based on the grades for all forms of classes conducted including lectures and laboratory works, self - preparation work and performance of individual tasks according to the program of the academic discipline “Statistical thinking for science about data”. Evaluation of the formed competences of students is conducted on the cumulative 100-point scale. According to the Provisional Regulations "On the assessment of students’ progress according to the cumulative rating system" at S. Kuznets KhNUE, the control methods include:

Current control over the semester during lectures, laboratory sessions and estimated by the sum of the points scored (the maximum amount is 60 points; the minimum amount that allows the student to take the exam - 35 points);

modular control carried out in the form of a colloquium as an intermediate mini-exam on the initiative of the teacher, taking into account the current control over the relevant content module and aims to integrate the evaluation of the student's learning outcomes after studying the material from the logically completed part of the discipline - content module;

final / semester control, conducted in the form of a semester exam, according to the schedule of the educational process.

The procedure for carrying out the current assessment of students' knowledge. Assessment of student's knowledge during seminars, practical and laboratory classes and performance of individual tasks is carried out according to the following criteria:

understanding, degree of assimilation of the theory and methodology of the problems under consideration; the degree of assimilation of the actual material of the discipline; acquaintance with the recommended literature, as well as contemporary literature on the issues under consideration; the ability to combine theory with practice when considering production situations, solving tasks, performing calculations in the process of performing individual tasks and tasks submitted for consideration in an audience; logic, structure, style of presentation of the material in written works and speeches in the audience, ability to substantiate their position, to generalize information and to draw conclusions; arithmetic correctness of the implementation of an individual and complex settlement task; the ability to conduct a critical and independent assessment of certain problem issues; the ability to explain alternative views and the presence of their own point of view, the position on a certain problematic issue; application of analytical approaches; quality and clarity of reasoning; logic, structuring and substantiation of conclusions on a specific problem; independence of work; literacy of presentation of the material; use of comparison methods, generalizations of concepts and phenomena; registration of work.

The general criteria for evaluating non-auditing independent work of students are: the depth and strength of knowledge, the level of thinking, the ability to systematize knowledge on specific topics, the ability to make sound conclusions, the possession of categorical apparatus, skills and techniques for the implementation of practical tasks, the ability to find the necessary information, carry out its systematization and processing, self-realization on practical and seminars.

The final control of the knowledge and competences of students in the discipline “Statistical thinking for science about data” is carried out on the basis of a semester examination, the task of which is to check the student's understanding of the program material in general, the logic and interrelations between the individual sections, the ability to use the accumulated knowledge creatively, the ability to formulate their attitude to a particular educational problem discipline, etc.

The examination ticket covers the program of discipline and involves determining the level of knowledge and the degree of mastery of competencies by students.

Each examination ticket consists of 5 practical situations (two stereotyped, two diagnostic and one heuristic task), which provide for the solution of typical professional tasks of the specialist in the workplace and allow to diagnose the level of theoretical training of the student and his level of competence in the discipline.

The result of the semester exam is evaluated in points (the maximum number is 40 points, the minimum number is scored - 25 points) and entered in the "Record of success" of the academic discipline.

The student should be considered certified if the sum of the points earned on the results of the final / semester test of success is equal to or exceeds 60. The minimum number of points for the current and modular control during the semester is 35 and the

minimum number of points scored on the exam is 25.

The final score in the discipline is calculated on the basis of the points obtained during the exam and the points obtained during the current control over the accumulation system. The total score in the points for the semester is: "60 and more points are counted", "59 and less points are not counted" and entered in the "Record of success" of the academic discipline.

Distribution of points for a week

(specify means of evaluation according to the technological card)

Name of content modules/themes			**Lectures**	**Laboratory sessions**	**Homeworks**	**The written test**	**Modular testing**	**Total**
Content module 1. Methodological bases of statistical modeling and forecasting	Topic 1	3, 4 Week	1	4	0,5	0,5		6
	Topic 2	5, 6 Week	1	4	0,5	0,5		6
	Topic 3	7 Week	0,5	4	0,5	0,5		5,5
	Topic 4	8 Week	0,5	4	0,5	0,5	10	15,5
Content module 2. Modeling and forecasting of multidimensional processes	Topic 5	9, 10 Week	1	4	0,5	0,5		6
	Topic 6	11 Week	0,5	4	0,5	0,5		5,5
	Topic 7	12 Week	0,5	4	0,5	0,5	10	15,5
	Examination							40
Total			**5**	**28**	**3,5**	**3,5**	**20**	**100**

Grading scale: national and ECTS

Assessment of the S. Kuznets KhNUE according to Economics scale	ECTS assessing scale		Assessment according to national scale
90-100	A	excellent performance	Excellent
82-89	B	above average	
74-81	C	work at all correct, but with a number of errors from	Good
64-73	D	not bad, but many drawbacks	Satisfactory
60-63	E	performance meets the minimum criteria	
35-59	FX	need to re-take	Unsatisfactory
1-34	F	repeat the discipline	

5. REFERENCES

Main

1. Антохонова И. В. Методы прогнозирования социально-экономических процессов: Учебное пособие. –Улан-Удэ: Изд-во ВСГТУ, 2004. – 212с.
2. Бабешко Л. О. Основы эконометрического моделирования: Учебное пособие. Изд. 3-е. - М.: Ком книга, 2007. - 432с.
3. Вітлинський В.В. Моделювання економіки: навч. посібн. / В. В. Вітлинський. – К.: КНЕУ, 2003. – 408 с.
4. Геєць В. М. Моделі і методи соціально-економічного прогнозування: підручник / В. М. Геєць, Т. С. Клебанова, О. І. Черняк, А. В. Ставицький та інші.-2 вид., виправ. -Х.: ВД «ІНЖЕК», 2008.- 396 с.
5. Єріна А. М. Статистичне моделювання та прогнозування: навч. посібн. / А. М. Єріна.– К.: КНЕУ, 2001.
6. Клебанова Т. С., Дубовина Н.А., Раевнева Е. В. Эконометрия: Учебно-методическое пособие для самостоятельного изучения дисциплины – Х.: Изд. Дом «ИНЖЭК», 2003. – 132 с.
7. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс: Учеб. — 8-е изд., испр. — М.: Дело, 2007. — 504 с.
8. Присенко Г.В. Прогнозування соціально-економічних процесів: Навч. посіб. / Г.В. Присенко, Є.І. Равікович — К.: КНЕУ, 2005. — 378 с.
9. Статистика. Навчальний посібник / Під ред. д.е.н., професора Раєвнєвої О.В. – Харків: Вид. ХНЕУ, 2010. – 520 с.
10. Сошникова Л. А., Тамашевич В. Н., Уебе Г., Шефер М. Многомерный статистический анализ в экономике: Учеб. Пособие для ВУЗов / Под ред. проф. В. Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.

Additional

11. Андрієнко В.Ю. Статистичні індекси в економічних дослідженнях / В. Ю. Андрієнко. – К. : 2004. – 536 с.
12. Дуброва Т.А., Павлов Д.Э., Осипова Н.П. Факторный анализ с использованием пакета “STATISTICA”. Учебное пособие / МГУ экономики, статистики и информатики, - М., 2002.
13. Клебанова Т.С., Забродский В.А, Полякова О.Ю., Петренко В.Л. Моделирование экономики. Учебное пособие. – Харьков : Изд. ХГЭУ, 2001. – 140 с
14. Организационно-экономическое моделирование : учебник : в 3 ч.. / А.И.Орлов. – М.: Изд-во МГТУ им. Н.Э. Баумана. – 2009. – 254 с.
15. Халафян А.А. STATISTICA 6. Статистический анализ данных. – М.: ООО «Бином-Пресс», 2008. – 512с.
16. Dickey D.A., Fuller W.A. Distribution of the estimators for autoregressive time-series with a unit root //Journal of the American statistical association. - 1979. - V. 74. - PP. 427-431
17. Granger C. W.J., Newbold P. Forecasting economic time series. 2nd ed. - N.Y.: Academic Press, 1986.- 324 p.
18. Lachenbruch, P.A. Discriminant Analysis. New York: Hafner., 1974 – 234 p.

Internet resources

19. Офіційний сайт державної служби статистики України [Електронний ресурс]. – Режим доступу: http://www.ukrstat.gov.ua/
20. Электронный учебник StatSoft [Електронний ресурс]. – Режим доступу : http: // www.statsoft.ru.