COURSE OUTLINE

(1) GENERAL

SCHOOL	Science and Technology
ACADEMIC UNIT	Science and Technology
PROGRAMME OF STUDIES	MSc in Data Science
LEVEL OF STUDIES	Postgraduate

COURSE CODE	DSC03	SEMESTER	1

COURSE TITLE	Statistical Methods for Data Science
COURSE TYPE *Elective, compulsory*	Compulsory
INSTRUCTOR(S)	Prof. Panayiotis Bozanis

INDEPENDENT TEACHING ACTIVITIES *if credits are awarded for separate components of the course, e.g. lectures, laboratory exercises, etc. If the credits are awarded for the whole of the course, give the weekly teaching hours and the total credits*	WEEKLY TEACHING HOURS	CREDITS
	3	6
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).		

TEACHING ACTIVITIES BREAKDOWN	WEEKLY HOURS
Theory	2,3
Recitation	0,7
Lab	0
Add rows if necessary. The organisation of teaching and the teaching methods used are described in detail at (d).	

COURSE TYPE *general background, special background, specialised general knowledge, skills development*	General Background
PREREQUISITE COURSES:	-
LANGUAGE OF INSTRUCTION and EXAMINATIONS:	English
IS THE COURSE OFFERED TO ERASMUS STUDENTS	Yes
COURSE WEBSITE (URL)	https://elearn-ucips.ihu.gr/

(2) LEARNING OUTCOMES

Learning outcomes
The course learning outcomes, specific knowledge, skills and competences of an appropriate level, which the students will acquire with the successful completion of the course are described. *Consult Appendix A* ● *Description of the level of learning outcomes for each qualifications cycle, according to the Qualifications Framework of the European Higher Education Area* ● *Descriptors for Levels 6, 7 & 8 of the European Qualifications Framework for Lifelong Learning and Appendix B* ● *Guidelines for writing Learning Outcomes*
On completing the course, the student will be able to: ● Understand the basic concepts of probability theory and statistics as they are applied in data science.

- Apply mathematical tools, models, and methods to data analysis tasks, such as data fitting, regression, sampling, hypothesis testing etc.
- Learn the fundamentals of statistical inference and its implementations.
- Use R to conduct for data analysis, processing and visualization.

General Competences

Taking into consideration the general competences that the degree-holder must acquire (as these appear in the Diploma Supplement and appear below), at which of the following does the course aim?

Search for, analysis and synthesis of data and information, with the use of the necessary technology	*Project planning and management*
	Respect for difference and multiculturalism
Adapting to new situations	*Respect for the natural environment*
Decision-making	*Showing social, professional and ethical responsibility and*
Working independently	*sensitivity to gender issues*
Team work	*Criticism and self-criticism*
Working in an international environment	*Production of free, creative and inductive thinking*
Working in an interdisciplinary environment	*……*
Production of new research ideas	*Others…*
	…….

- Search for, analysis and synthesis of data and information, with the use of the necessary technology
- Decision Making
- Working independently
- Production of free, creative, and inductive thinking

(3) SYLLABUS

The course overviews basic statistical foundations of Data Science and presents the most commonly used statistical methods in the field. The students will gain the necessary conceptual understanding of statistical methods used to analyze and interpret massive data sets as well as extract meaningful conclusions out of them. In addition, they will be able to apply mathematical tools, models and methods to data analysis tasks, such as data fitting, regression, sampling, hypothesis testing etc. using R. The topics covered include:

- Descriptive Statistics.
- Probability Distributions.
- Sampling and Sampling Distributions.
- Interval Estimation.
- Hypothesis Testing.
- Statistical Inference.
- ANOVA.
- Linear Regression.
- Nonparametric Methods.

(4) TEACHING and LEARNING METHODS - EVALUATION

| DELIVERY
Face-to-face, Distance learning, etc.	Hybrid: Face to face and synchronous distance learning
USE OF INFORMATION AND COMMUNICATIONS TECHNOLOGY
Use of ICT in teaching, laboratory education, communication with students	**Use of ICT in Teaching**
During the educational process, various machine learning and programming tools are used, along with the material available at the e-learning platform.
The hybrid teaching method involves synchronous learning with the support of the videoconferencing tool Zoom.
Students are taught a variety of tools related to the course content and material.

Use of ICT in Communication with students |

	• The course material (slides, scientific articles, exercises, etc.) is posted on the course page at the e-learn platform (Moodle). • Use of Moodle Forums announcements. • Live video meetings via Zoom/Teams. • Contact via email.

TEACHING METHODS	Activity	Semester workload
The manner and methods of teaching are described in detail.	Lectures	30 hrs.
Lectures, recitation, seminars, laboratory practice, fieldwork, study and analysis of bibliography, tutorials, placements, clinical practice, art workshop, interactive teaching, educational visits, project, essay writing, artistic creativity, etc.	Recitation	9 hrs.
	Homework	18 hrs.
	Non-Directed Study	93 hrs.
The student's study hours for each learning activity are given as well as the hours of non-directed study according to the principles of the ECTS		
	Course total	**150 hrs.**

COURSE MATERIAL ARRANGEMENT	Theory/Recitation	
	Introduction to R	4 hrs.
	Introduction to Data and Statistics, Descriptive Statistics	5 hrs.
	Sampling and Sampling Distributions	3 hrs.
	Interval Estimation	3 hrs.
	Hypothesis Test	3 hrs.
	Statistical Inference about means and proportions with two populations	2 hrs.
	Inference about populations variances, Tests of goodness of fit and independence	2.5 hrs.
	Experimental Design and ANOVA	3.5 hrs.
	Simple Linear Regression	3.5 hrs.
	Multiple Linear Regression	3.5 hrs.
	Regression Analysis: Model Building	3 hrs.
	Nonparametric Methods	3 hrs.

STUDENT PERFORMANCE EVALUATION	

Description of the evaluation procedure
 	
Language of evaluation, methods of evaluation, summative or conclusive, multiple choice questionnaires, short-answer questions, open-ended questions, problem solving, written work, essay/report, oral examination, public presentation, laboratory work, clinical examination of patient, art interpretation, other
 	
Specifically-defined evaluation criteria are given, and if and where they are accessible to students
	Language of Evaluation: English
	
	Evaluation Procedure:
	● Homework (100%).
	    ○ Use R to solve various problems covering the course topics.
	    ○ The homework should be completed individually.
	The evaluation procedure is announced to the students during the first lecture and is also accessible at the e-learn platform throughout the entire semester.

STUDENT OBLIGATIONS	
Compulsory attendance of lectures, labs, recitations, compulsory participation in midterms, exams, compulsory delivery of homework, projects, etc.	● Compulsory attendance of lectures ● Compulsory attendance of recitation ● Compulsory delivery of homework

(5) ATTACHED BIBLIOGRAPHY

- Suggested Textbooks

1. Anderson D.R., Sweeney D.J., Williams T.A., Camm J.D., Cochran J.J, Fry M.J., Ohlmann, J.W., Statistics for Business & Economics, Cengate, 14th edition, 2020.
2. Stinerock R., Statistics with R: A Beginner's Guide, Sage Publishing, 2018.

- Additional Bibliography:

1. Heumann C., Schomaker M., Shalabh, Introduction to Statistics and Data Analysis: With Exercises, Solutions and Applications in R, Springer, 2016.
2. McClave J.T., Benson P.J., Sincich T., Statistics for Business & Economics, Pearson, 13th edition, 2018.