

**FAC.SIMILE SCHEDA DI INSEGNAMENTO CON LE INDICAZIONI SPECIFICHE PER CIASCUN CAMPO DA COMPILARE (DA PREDISPORRE SU CARTA INTESTATA DEL DIPARTIMENTO/SCUOLA)**

**CORSO DI STUDIO: Laurea Magistrale in Bioinformatica (LM-8)**

**ANNO ACCADEMICO: 2023-2024**

**DENOMINAZIONE DELL'INSEGNAMENTO: Fondamenti di Matematica per l'Analisi dei Dati (integrato con Metodi Numerici per la Bioinformatica)**

<b>Principali informazioni sull'insegnamento</b>	
Anno di corso	<i>1 anno</i>
Periodo di erogazione	<i>1 semestre (02.10.2023 19.01.2024)</i>
Crediti formativi universitari (CFU/ETCS):	<i>3 CFU</i>
SSD	<i>MAT/08</i>
Lingua di erogazione	<i>Italiano</i>
Modalità di frequenza	<i>Facoltativa (fortemente consigliata)</i>

<b>Docente</b>	
Nome e cognome	<i>Flavia Esposito</i>
Indirizzo mail	<i>flavia.esposito@uniba.it</i>
Telefono	<i>0805442711</i>
Sede	<i>Stanza n 24, Il piano, Dipartimento di Matematica</i>
Sede virtuale	<i>cuzrciw</i>
Ricevimento	<i>Su appuntamento, da concordare per e-mail; in presenza o in remoto</i>

<b>Organizzazione della didattica</b>			
<b>Ore</b>			
Totali	Didattica frontale	Pratica (laboratorio, campo, esercitazione, altro)	Studio individuale
<i>75</i>	<i>16</i>	<i>15</i>	<i>44</i>
<b>CFU/ETCS</b>			
<i>3</i>	<i>2</i>	<i>1</i>	

<b>Obiettivi formativi</b>	<i>Acquisizione degli strumenti matematico-statistici per l'analisi dei dati e delle tecniche numeriche per il pre-processing del dato</i>
<b>Prerequisiti</b>	<i>Conoscenze di base di matematica coerenti con quelle richieste per l'ammissione al corso di laurea</i>
<b>Metodi didattici</b>	<i>- Didattica frontale condotta con l'ausilio di supporti didattici (slide). - Esercitazioni al calcolatore con il supporto di codici in linguaggi open source forniti dal docente e utilizzo di dati di benchmark presenti nei repository online</i>

<p><b>Risultati di apprendimento previsti</b></p>	<p><i>Al termine dell'insegnamento lo studente dovrà essere in grado di applicare le tecniche matematico statistiche di analisi di dati di tipo biomedico e identificare le eventuali problematiche da trattare</i></p>
<p><b>DD1 Conoscenza e capacità di comprensione</b></p>	<p><b>DD1:</b> <i>conoscenza e capacità di comprensione;</i></p> <ul style="list-style-type: none"> <li>• <i>struttura e tipologia dei dati</i></li> <li>• <i>Concetti di statistica descrittiva per l'analisi del dato</i></li> <li>• <i>Test di ipotesi di statistica parametrica e non parametrica</i></li> <li>• <i>Approcci e tecniche matematiche per il pre-processing del dato</i></li> <li>• <i>Codici numerici che implementano le tecniche standard per l'analisi dei dati</i></li> </ul>
<p><b>DD2 Conoscenza e capacità di comprensione applicate</b></p>	<p><b>DD2:</b> <i>capacità di applicare conoscenza e comprensione;</i></p> <ul style="list-style-type: none"> <li>• <i>Conoscenza dei concetti matematico-statistici e utilizzo delle tecniche standard per l'analisi dei dati biomedici</i></li> <li>• <i>Implementazione di workflow per l'analisi dei dati biomedici</i></li> </ul>
<p><b>DD3-5 Competenze trasversali</b></p>	<p><b>DD3:</b> <i>capacità critiche e di giudizio</i></p> <p><b>Autonomia di giudizio</b></p> <p><i>Al termine dell'insegnamento lo studente dovrà essere in grado di</i></p> <ul style="list-style-type: none"> <li>• <i>Comprendere la struttura del dato proposto e identificare la metodologia più appropriata per descriverlo</i></li> <li>• <i>Individuare la tecnica adatta a rispondere ai quesiti posti</i></li> <li>• <i>Gestire la presenza di outlier, missing data e saper gestire eventuali normalizzazioni</i></li> <li>• <i>Implementare un workflow per l'analisi del dato</i></li> </ul> <p><b>DD4:</b> <i>capacità di comunicare quanto si è appreso.</i></p> <ul style="list-style-type: none"> <li>• <i>Abilità comunicative</i></li> </ul> <p><i>Al termine dell'insegnamento lo studente dovrà essere in grado di</i></p> <ul style="list-style-type: none"> <li>• <i>Esporre i concetti appresi e le scelte fatte per l'analisi del dato in esame con formalismo matematico</i></li> </ul> <p><b>DD5:</b> <i>capacità di proseguire lo studio in modo autonomo nel corso della vita..</i></p> <ul style="list-style-type: none"> <li>• <i>Capacità di apprendere in modo autonomo</i></li> </ul> <p><i>Al termine dell'insegnamento lo studente dovrà essere in grado di</i></p> <ul style="list-style-type: none"> <li>• <i>Dimostrare di aver acquisito un metodo di studio adeguato, supportato dalla consultazione dei testi, da percorsi di approfondimento individuali o e in gruppo e dalla implementazione al calcolatore delle tecniche esposte durante il corso</i></li> </ul>

<b>Contenuti di insegnamento (Programma)</b>	<p><i>-Statistica descrittiva per l'analisi esplorativa dei dati. Tabelle di contingenza e rappresentazioni grafiche. Definizione di campione aleatorio, di statistica e di stimatore. Esempi di stimatori della media e della varianza di una variabile aleatoria. Test parametrici di verifica delle ipotesi. Ipotesi semplici e ipotesi composte. L'errore del primo e del secondo tipo. Livello di significatività di un test. I test per i parametri della distribuzione normale. Test per la media, per il confronto di medie tra campioni indipendenti e dipendenti, per il confronto di varianze. L'analisi per grandi campioni.</i></p> <p><i>-Statistica non parametrica. Il test del chi-quadro di bontà di adattamento, di indipendenza, di omogeneità. Il test dei ranghi di Wilcoxon. Il test dei ranghi di Wilcoxon-Mann-Whitney per due campioni indipendenti. Il test di Wilcoxon dei ranghi segnati (per dati appaiati). Il test di bontà di adattamento di Kolmogorov-Smirnov.</i></p> <p><i>-Analisi della Varianza a una via. Cenni di analisi della varianza a due vie. Il test per i confronti multipli di Tukey. Correlazione, coefficienti di correlazione di Pearson e Spearman.</i></p> <p><i>-Pre-processing del dato: definizione/classificazione e trattamento di missing values, definizione outlier e identificazione outlier. Cenni sui metodi di individuazione e trattamento statistico di outlier. Trasformazioni e normalizzazioni del dato. Cenni sul problema della Curse of Dimensionality e come risolverlo con tecniche di fattorizzazione lineare.</i></p> <p><i>-Esercitazioni al calcolatore tramite il software statistico R <a href="https://www.r-project.org">https://www.r-project.org</a></i></p>
<b>Testi di riferimento</b>	<p><i>-Carl D. Meyer - Matrix Analysis and Applied Linear Algebra: 1st (first). Siam, 2001. <a href="http://matrixanalysis.com/">http://matrixanalysis.com/</a></i></p> <p><i>-Gilbert Strang - Linear Algebra and Learning from Data, Wellesley-Cambridge Press, 2019 <a href="https://math.mit.edu/~gs/learningfromdata/">https://math.mit.edu/~gs/learningfromdata/</a></i></p> <p><i>-Robert Gentleman, Vince Carey, Wolfgang Huber, Rafael Irizarry, Sandrine Dudoit, Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer 2005. (<a href="http://www.bioconductor.org/help/publications/books/bioinformatics-and-computational-biology-solutions/">http://www.bioconductor.org/help/publications/books/bioinformatics-and-computational-biology-solutions/</a>)</i></p> <p><i>-Matthew He, Sergey Petoukhov. Mathematics of Bioinformatics: Theory, Practice, and Applications. Wiley&amp;Sons 2011 <a href="https://onlinelibrary.wiley.com/doi/book/10.1002/9780470904640">https://onlinelibrary.wiley.com/doi/book/10.1002/9780470904640</a></i></p> <p><i>- Michael C. Whitlock, Dolph Schluter. Analisi statistica dei dati biologici. Zanichelli</i></p> <p><i>-I.T. Jolliffe, Principal Component Analysis, Second Edition, 2002, Springer <a href="https://link.springer.com/book/10.1007/b98835">https://link.springer.com/book/10.1007/b98835</a></i></p>
<b>Note ai testi di riferimento</b>	<p><i>Appunti e riferimenti specifici forniti dal docente</i></p>
<b>Materiali didattici</b>	<p><i>Il materiale didattico sarà reso disponibile sulla classe Teams dell'insegnamento</i></p>

<b>Valutazione</b>	
<b>Modalità di verifica dell'apprendimento</b>	<p><i>Prova Orale sul programma svolto nel corso delle lezioni ed esercitazioni e/o progetto assegnato dal docente</i></p>
<b>Criteri di valutazione</b>	<ul style="list-style-type: none"> <li>• <i>Conoscenza e capacità di comprensione:</i> <ul style="list-style-type: none"> <li>○ <i>Lo studente deve dimostrare una adeguata conoscenza dei contenuti dell'insegnamento</i></li> </ul> </li> <li>• <i>Conoscenza e capacità di comprensione applicate:</i> <ul style="list-style-type: none"> <li>○ <i>Lo studente deve dimostrare una adeguata conoscenza delle possibili applicazioni dei concetti teorici e possedere una adeguata capacità di implementare tali applicazioni</i></li> </ul> </li> <li>• <i>Autonomia di giudizio:</i></li> </ul>

	<ul style="list-style-type: none"> <li>○ <i>Lo studente deve dimostrare una adeguata autonomia nella selezione dei concetti teorici più idonei alla risoluzione di problemi pratici</i></li> <li>• <i>Abilità comunicative:</i> <ul style="list-style-type: none"> <li>○ <i>Lo studente deve dimostrare una adeguata capacità espositiva dei contenuti studiati e una adeguata capacità di analisi e sintesi</i></li> </ul> </li> <li>• <i>Capacità di apprendere:</i> <ul style="list-style-type: none"> <li>○ <i>Lo studente deve dimostrare una buona capacità di effettuare collegamenti interdisciplinari</i></li> </ul> </li> </ul>
<p>Criteria di misurazione dell'apprendimento e di attribuzione del voto finale</p>	<p><i>Il voto finale è attribuito in trentesimi. L'esame si intende superato quando il voto è maggiore o uguale a 18. Per la formulazione del voto finale si prenderanno in considerazione i seguenti indicatori: grado di conoscenza dei contenuti e degli argomenti dell'insegnamento, capacità e correttezza nell'applicare i concetti fondamentali trattati durante le lezioni frontali e le esercitazioni, qualità della esposizione orale.</i></p> <p><i>Tutti gli argomenti del programma contribuiscono in modo uguale alla formulazione del voto finale.</i></p>
<p><b>Altro</b></p>	

**COURSE OF STUDY: Master Course in Bioinformatics (LM-8)**
**ACADEMIC YEAR: 2023-2024**
**ACADEMIC SUBJECT: Mathematical Foundation of Data Analysis (integrated with Numerical Methods for Bioinformatics)**

General information	
Year of the course	First
Academic calendar (starting and ending date)	First Semester (02.10.2023-19.01.2024)
Credits (CFU/ETCS):	3
SSD	MAT/08
Language	Italian
Mode of attendance	Teaching attendance strongly recommended

Professor/ Lecturer	
Name and Surname	<i>Flavia Esposito</i>
E-mail	<i>flavia.esposito@uniba.it</i>
Telephone	<i>0805442711</i>
Department and address	<i>Room 24, II floor, Department of Mathematics</i>
Virtual room	<i>czrciw</i>
Office Hours (and modalities: e.g., by appointment, online, etc.)	<i>By appointment, to be arranged by e-mail; in-person or remote</i>

Work schedule			
Hours			
Total	Lectures	Hands-on (laboratory, workshops, working groups, seminars, field trips)	Out-of-class study hours/ Self-study hours
<i>75</i>	<i>16</i>	<i>15</i>	<i>44</i>
CFU/ETCS			
<i>3</i>	<i>2</i>	<i>1</i>	

<b>Learning Objectives</b>	Acquisition of classical numerical methodologies and linear algebra techniques for the analysis of structured biomedical data.
<b>Course prerequisites</b>	Basic knowledge of mathematics consistent with that required for admission to the degree program

<b>Teaching strategie</b>	- Frontal teaching conducted with the aid of slides. - Computer exercises with the support of codes in open-source languages provided by the lecturer and use of benchmark data stored in online repositories
<b>Expected learning outcomes in terms of</b>	
<b>Knowledge and understanding on:</b>	<ul style="list-style-type: none"> <li>○ Structures and types of data</li> <li>○ Descriptive statistics concepts for data analysis</li> <li>○ Hypothesis testing of parametric and non-parametric statistics</li> <li>○ Mathematical approaches and techniques for data pre-processing</li> <li>○ Numerical codes implementing standard techniques for data analysis</li> </ul>
<b>Applying knowledge and understanding on:</b>	<ul style="list-style-type: none"> <li>○ Knowledge of mathematical and statistical concepts and use of standard techniques for the analysis of biomedical data</li> <li>○ Implementation of workflows for the analysis of biomedical data</li> </ul>
<b>Soft skills</b>	Making informed judgments and choices <ul style="list-style-type: none"> <li>○ Understanding the structure of the proposed data and identifying the most appropriate methodology to describe it</li> </ul>

	<ul style="list-style-type: none"> <li>○ Identifying the appropriate technique to answer the questions posed</li> <li>○ Managing the presence of outliers, missing data and know how to handle possible normalization</li> <li>○ Implementing a workflow for data analysis</li> </ul> <p>Communicating knowledge and understanding</p> <ul style="list-style-type: none"> <li>○ Explaining the concepts learned and the choices made for the analysis of the data under examination using mathematical formalism</li> </ul> <p>Capacities to continue learning</p> <ul style="list-style-type: none"> <li>○ Demonstrating to have acquired an adequate study method, supported by the consultation of texts, individual or group in-depth studies and computer implementation of the techniques presented during the course.</li> </ul>
--	---

<b>Syllabus</b>	
<b>Content knowledge</b>	<p>-Descriptive statistics for exploratory data analysis. Contingency tables and graphical representations. Definition of a random sample, statistic, and estimator. Examples of estimators of the mean and variance of a random variable. Parametric tests of hypothesis testing. Simple and compound hypotheses. The error of the first and second type. The significance level of a test. Tests for the parameters of the normal distribution. Tests for the mean, for comparison of means between independent and dependent samples, for comparison of variances. The analysis for large samples.</p> <p>-Non-parametric statistics. The chi-square test for goodness of fit, independence, and homogeneity. The Wilcoxon ranks test. The Wilcoxon-Mann-Whitney rank test for two independent samples. The Wilcoxon test of signed ranks (for paired data). The Kolmogorov-Smirnov goodness-of-fit test.</p> <p>-One-way analysis of variance. Outlines of two-way analysis of variance. Tukey's multiple comparisons test. Correlation, Pearson, and Spearman correlation coefficients.</p> <p>-Data pre-processing: definition/classification and treatment of missing values, outlier definition, and outlier identification. Hints on outlier identification and statistical treatment methods. Transformations and normalizations of the data. Notes on the Curse of Dimensionality problem and how to solve it using linear factorization techniques.</p> <p>-Exercises using the statistical software R <a href="https://www.r-project.org">https://www.r-project.org</a></p>
<b>Texts and readings</b>	<p>Carl D. Meyer - Matrix Analysis and Applied Linear Algebra: 1st (first). Siam, 2001. <a href="http://matrixanalysis.com/">http://matrixanalysis.com/</a></p> <p>-Gilbert Strang - Linear Algebra and Learning from Data, Wellesley-Cambridge Press, 2019 <a href="https://math.mit.edu/~gs/learningfromdata/">https://math.mit.edu/~gs/learningfromdata/</a></p> <p>-Robert Gentleman, Vince Carey, Wolfgang Huber, Rafael Irizarry, Sandrine Dudoit, Bioinformatics and Computational Biology Solutions Using R and Bioconductor, Springer 2005. (<a href="http://www.bioconductor.org/help/publications/books/bioinformatics-and-computational-biology-solutions/">http://www.bioconductor.org/help/publications/books/bioinformatics-and-computational-biology-solutions/</a>)</p> <p>-Matthew He, Sergey Petoukhov. Mathematics of Bioinformatics: Theory, Practice, and Applications. Wiley&amp;Sons 2011 <a href="https://onlinelibrary.wiley.com/doi/book/10.1002/9780470904640">https://onlinelibrary.wiley.com/doi/book/10.1002/9780470904640</a></p> <p>- Michael C. Whitlock, Dolph Schluter. <i>Analisi statistica dei dati biologici</i>. Zanichelli</p> <p>-I.T. Jolliffe, Principal Component Analysis, Second Edition, 2002, Springer <a href="https://link.springer.com/book/10.1007/b98835">https://link.springer.com/book/10.1007/b98835</a></p>
<b>Notes, additional materials</b>	Notes and slides provided by the instructor
<b>Repository</b>	Teaching materials will be made available on the MTeams class of the course

<b>Assessment</b>	
Assessment methods	<i>Oral examination on the syllabus and exercises and/or project assigned by the lecturer</i>
Assessment criteria	<ul style="list-style-type: none"> <li>• <i>Knowledge and understanding</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate adequate knowledge of the main topics of the course</i></li> </ul> </li> <li>• <i>Applying knowledge and understanding</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate adequate knowledge of the possible applications of the theoretical concepts and possess adequate ability to implement these applications</i></li> </ul> </li> <li>• <i>Autonomy of judgment</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate adequate autonomy in selecting the most appropriate theoretical concepts for solving practical problems.</i></li> </ul> </li> <li>• <i>Communicating knowledge and understanding</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate an adequate expository capacity of the studied topic</i></li> </ul> </li> <li>• <i>Communication skills</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate an adequate capacity in analysis and synthesis</i></li> </ul> </li> <li>• <i>Capacities to continue learning</i> <ul style="list-style-type: none"> <li>• <i>Students must demonstrate a good ability to make interdisciplinary connections</i></li> </ul> </li> </ul>
Final exam and grading criteria	The final grade is given in thirtieths. The exam is considered passed when the grade is greater than or equal to 18. The following indicators will be considered in formulating the final grade: degree of knowledge of the content and topics of the teaching, ability, and correctness in applying the fundamental concepts covered during the lectures and exercises, quality of oral exposition. All program topics contribute equally to the formulation of the final grade.
<b>Further information</b>	
	.