

---

# Testi del Syllabus

---

Resp. Did.	<b>PIAZZA SILVANO</b>	<b>Matricola: 041786</b>
Docente	<b>PIAZZA SILVANO, 3 CFU</b>	
Anno offerta:	<b>2023/2024</b>	
Insegnamento:	<b>919SV - ANALISI DI DATI SPERIMENTALI E BIOINFORMATICI</b>	
Corso di studio:	<b>SM53 - GENOMICA FUNZIONALE</b>	
Anno regolamento:	<b>2023</b>	
CFU:	<b>3</b>	
Settore:	<b>MED/01</b>	
Tipo Attività:	<b>D - A scelta dello studente</b>	
Anno corso:	<b>1</b>	
Periodo:	<b>Secondo Semestre</b>	
Sede:	<b>TRIESTE</b>	

---

## Testi in italiano

**Lingua insegnamento** ITALIANO

**Contenuti (Dipl.Sup.)** Il corso descrive come capire e ottimizzare la raccolta dei dati sperimentali, ottenuti tramite tecniche come sequenziamento del DNA, esperimenti di espressione genica, proteomica, trascrittomica, metabolomica, etc. Questi dati possono essere sia grezzi (raw data) che pre-elaborati. Verranno trattati i seguenti argomenti:

- Pre-elaborazione dei dati: i dati grezzi vengono elaborati per rimuovere eventuali effetti confondenti o errori sperimentali.
- Analisi statistica: i dati pre-elaborati vengono sottoposti ad analisi statistica per identificare i modelli, le relazioni e le differenze significative.
- Identificazione delle feature significative: identificare le feature o le variabili biologiche che sono significativamente associate a un fenomeno o condizione.
- Interpretazione dei risultati: interpretare i risultati delle analisi alla luce delle domande di ricerca o degli obiettivi dell'esperimento.
- Visualizzazione dei dati: l'utilizzo di grafici e visualizzazioni è un aspetto importante dell'analisi dei dati sperimentali e bioinformatici.

**Testi di riferimento** Per la preparazione dell'esame gli studenti frequentanti potranno utilizzare gli appunti personali e fare riferimento alle presentazioni PDF delle lezioni, disponibili sulla piattaforma Moodle2/Teams del corso. Gli studenti (in particolare se non frequentanti) possono inoltre fare riferimento alle seguenti fonti bibliografiche utilizzate per preparare le lezioni:

Learning Statistics with R (<https://learningstatisticswithr.com/>)  
Modern Statistics with R (<https://www.modernstatisticswithr.com/>)  
simpleR - Using R for Introductory Statistics (<https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>)

<b>Obiettivi formativi</b>	<p>In accordo con i principi "Descrittori di Dublino" per i Corsi di Studio Magistrali, il corso ha l'obiettivo di consentire agli Studenti di dimostrare:</p> <p>D1. Conoscenza e comprensione.  Un obiettivo chiave sarà sviluppare le competenze per l'analisi bioinformatica dei dati biologici, tramite l'apprendimento di strumenti e software specifici per l'elaborazione, l'analisi statistica e l'interpretazione dei dati bioinformatici. Imparare le tecniche di pre-elaborazione dei dati per rimuovere rumori, correggere errori o normalizzare i dati. Acquisire concetti statistici di base, quali i test ipotetici, l'analisi della varianza, le correlazioni, la regressione per l'analisi dei dati biologici.</p> <p>D2. Capacità di applicare conoscenza e comprensione  Gli studenti saranno in grado di applicare tecniche e strumenti bioinformatici per analizzare dati biologici, elaborare e visualizzare risultati. Inoltre gli studenti svilupperanno competenze nella gestione e nell'analisi di grandi dataset omici (genomica, trascrittomica, proteomica, etc.)</p> <p>D3. Autonomia di giudizio.  Gli studenti svilupperanno la capacità di valutare criticamente i risultati delle analisi bioinformatiche e di trarre conclusioni basate sui dati. Inoltre svilupperanno la capacità di selezionare e adattare metodi bioinformatici appropriati per risolvere problemi di ricerca specifici. La capacità di applicare conoscenza e l'autonomia di giudizio verranno sviluppate in particolare durante lo svolgimento di progetti durante i quali gli studenti apprenderanno criteri per scegliere strategie sperimentali opportune e impareranno a interpretare i dati sperimentali ottenuti.</p> <p>D4. Abilità comunicative.  Gli studenti impareranno a comunicare in modo chiaro e conciso i risultati dell'analisi dei dati sperimentali e bioinformatici. Ciò includerà la preparazione di report scientifici, la presentazione di risultati in forma orale e la visualizzazione dei dati tramite grafici e figure. Saranno inoltre incentivati a porre domande durante le lezioni per migliorare il lessico scientifico.</p> <p>D5 Capacità di apprendimento.  Gli studenti impareranno a sviluppare le competenze necessarie per intraprendere nuovi metodi ed analisi per gli studi successivi con un alto grado di autonomia.</p>
<b>Prerequisiti</b>	Non sono previste propedeuticità.
<b>Metodi didattici</b>	Il corso viene svolto attraverso lezioni frontali convenzionali incentivando gli studenti ad interagire proponendo soluzioni in risposta alle domande poste dal docente. Lo svolgimento delle lezioni comprenderà l'utilizzo di software dedicato (R/bioconductor).
<b>Altre informazioni</b>	Sulle piattaforme Moodle/Teams del corso sono disponibili: il programma dettagliato e il calendario delle lezioni, le diapositive delle lezioni e materiale didattico di supporto.
<b>Modalità di verifica dell'apprendimento</b>	<p>Durante il corso verranno distribuiti dei progetti da svolgere in gruppo, attraverso una prova pratica di analisi di dati utilizzando il software R, i cui risultati saranno esposti a fine corso. La sessione di esame prevede inoltre lo svolgimento di una prova scritta composta da domande aperte sull'intero programma.</p> <p>La valutazione sarà basata per l'80% sul risultato del progetto e per il 20% sul risultato della prova scritta (20%).</p> <p>La griglia di valutazione adottata è la seguente:</p> <ul style="list-style-type: none"> <li>- Eccellente (30 - 30 e lode): ottima conoscenza degli argomenti, proprietà di linguaggio e capacità di approfondimento.</li> <li>- Molto buono (27 - 29): buona conoscenza degli argomenti, notevole proprietà di linguaggio, buona capacità di approfondimento.</li> <li>- Buono (24-26): buona conoscenza dei principali argomenti, discreta proprietà di linguaggio e capacità di approfondimento.</li> <li>- Soddisfacente (21-23): lo/la studente/essa non mostra piena padronanza degli argomenti principali dell'insegnamento, pur possedendone le conoscenze fondamentali; mostra comunque soddisfacente proprietà di linguaggio.</li> <li>- Sufficiente (18-20): minima conoscenza degli argomenti principali</li> </ul>

dell'insegnamento e del linguaggio tecnico.

- Insufficiente: lo/la studente/essa non possiede una conoscenza accettabile dei contenuti dei diversi argomenti del programma. Gli eventuali studenti non frequentanti che non potessero partecipare al progetto potranno preparare un breve saggio per approfondire un argomento specifico previamente concordato con il docente.

## Programma esteso

Il corso descrive come capire e ottimizzare la raccolta dei dati sperimentali, che possono essere ottenuti tramite tecniche come sequenziamento del DNA, esperimenti di espressione genica, proteomica, trascrittomica, metabolomica, etc. Questi dati possono essere sia grezzi (raw data) che pre-elaborati. Verranno trattati i seguenti argomenti:

- Pre-elaborazione dei dati: i dati grezzi vengono elaborati per rimuovere eventuali effetti confondenti o errori sperimentali. Questo può includere la correzione degli errori di sequenziamento del DNA, la normalizzazione dei dati di espressione genica, la rimozione dei dati di bassa qualità o la gestione dei valori mancanti.

- Analisi statistica: Una volta pre-elaborati, i dati vengono sottoposti ad analisi statistica per identificare i modelli, le relazioni o le differenze significative. Ciò può comportare l'utilizzo di test statistici, come il t-test o l'analisi della varianza (ANOVA), per confrontare gruppi di campioni o il calcolo delle correlazioni tra diverse variabili.

- Identificazione delle feature significative: In molti casi, è necessario identificare le feature o le variabili biologiche che sono significativamente associate a un determinato fenomeno o condizione. Questo può essere fatto utilizzando metodi di selezione delle feature, come l'analisi delle componenti principali (PCA), la regressione logistica o l'albero di decisione.

- Interpretazione dei risultati: Una volta completate le analisi, è importante interpretare i risultati alla luce delle domande di ricerca o degli obiettivi dell'esperimento. Questo può richiedere l'uso di database biologici, strumenti bioinformatici o la consultazione della letteratura scientifica per ottenere una migliore comprensione delle implicazioni biologiche dei risultati.

- Visualizzazione dei dati: L'uso di grafici e visualizzazioni è un aspetto importante dell'analisi dei dati sperimentali e bioinformatici. Le visualizzazioni possono aiutare a identificare pattern o tendenze nei dati, rendendo più semplice la comunicazione dei risultati in modo chiaro e comprensibile.

## Obiettivi Agenda 2030 per lo sviluppo sostenibile

Questo insegnamento approfondisce argomenti strettamente connessi a uno o più obiettivi dell'Agenda 2030 per lo Sviluppo Sostenibile delle Nazioni Unite.

## Obiettivi per lo sviluppo sostenibile

Codice	Descrizione
3	Salute e benessere



## Testi in inglese

	Italian
--	---------

The course describes how to understand and optimise the collection of experimental data, obtained by techniques such as DNA sequencing, gene expression experiments, proteomics, transcriptomics, metabolomics, etc. These data can be either raw data or pre-processed. The following topics will be covered:

- Data pre-processing: raw data are processed to remove any confounding effects or experimental errors.

- Statistical analysis: once pre-processed, data are subjected to statistical analysis to identify significant patterns, relationships or differences.

- Identification of significant features: to identify features or biological variables that are significantly associated with a given phenomenon or condition.
- Interpretation of results: once the analyses have been completed, it is important to interpret the results in light of the research questions or objectives of the experiment.
- Visualisation of data: the use of graphs and visualisations is an important aspect of experimental and bioinformatics data analysis.

For exam preparation, attending students may use their personal notes and refer to the PDF presentations of the lectures, available on the course's Moodle2/Teams platforms. Students (particularly non-attending students) may also refer to the following bibliographic sources used to prepare for the lectures:

Learning Statistics with R (<https://learningstatisticswithr.com/>)  
 Modern Statistics with R (<https://www.modernstatisticswithr.com/>)  
 simpleR - Using R for Introductory Statistics (<https://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>)

In accordance with the Dublin Descriptors for Master's Degree Courses, the course aims to enable students to demonstrate:

D1. Knowledge and Understanding.

A key objective will be to develop skills for the bioinformatic analysis of biological data. This includes learning specific tools and software used for the processing, statistical analysis and interpretation of bioinformatics data. Learning data pre-processing techniques to remove noise, correct errors or normalise data. Students will learn common techniques and methods for data pre-processing. Acquire basic statistical concepts such as hypothesis testing, analysis of variance, correlations, regression for the analysis of biological data.

D2. Applying knowledge and understanding

Students will be able to apply bioinformatics techniques and tools to analyse biological data, process and visualise results. Furthermore, students will develop skills in the management and analysis of large omics datasets (genomics, transcriptomics, proteomics, etc.).

D3. Making judgements.

Students will develop the ability to critically evaluate the results of bioinformatic analyses and to draw conclusions based on the data. In addition, they will develop the ability to select and adapt appropriate bioinformatic methods to solve specific research problems. The ability to apply knowledge and autonomy of judgement will be developed in particular during the course projects where students will learn criteria for choosing appropriate experimental strategies and to interpret the experimental data obtained

D4. Communication skills.

Students will learn how to communicate in a clear and concise way the results of experimental and bioinformatics data analysis. This will include: preparation of scientific reports, presentation of results in oral form and visualisation of data by means of graphs and figures. Students will also be encouraged to ask questions and improve their scientific vocabulary.

D5 Learning skills.

Students will develop skills necessary to undertake new methods and analyses for further study with a high degree of autonomy.

There are no propaedeutic courses.

The course is conducted through conventional lectures, encouraging students to interact by proposing solutions to questions posed by the lecturer. The lectures will include the use of dedicated software (R/bioconductor).

Students have access to the slides of the lectures through the Moodle/Teams platform, together with other teaching material.

The course includes group work (projects of data analysis using R software) to be presented at the end of the course. A final exam will consist of a written test with open questions on the entire programme. Evaluation will be based for 80% on the project carried out and its presentation and for 20% on the written test.

The evaluation grid adopted is as follows:

- Excellent (30 - 30 cum laude): very good knowledge of the topics, property of language and capacity for in-depth study.
- Very good (27 - 29): good knowledge of the topics, remarkable ownership of language, good capacity for in-depth study.
- Good (24-26): good knowledge of the main topics, fair properties of language and ability to deepen.
- Satisfactory (21-23): the student does not show full mastery of the main topics of the course, even though he/she possesses the fundamental knowledge; however, he/she shows satisfactory properties of language.
- Sufficient (18-20): minimal knowledge of the main teaching topics and technical language.
- Insufficient: the student does not possess an acceptable knowledge of the content of the various topics in the syllabus.

Non-attending students may prepare a short essay on a specific topic agreed in advance with the lecturer.

The course describes how to understand and optimise the collection of experimental data, which can be obtained by techniques such as DNA sequencing, gene expression experiments, proteomics, transcriptomics, metabolomics, etc. These data can be either raw data or pre-processed. The following topics will be covered:

- Data pre-processing: raw data are processed to remove any confounding effects or experimental errors. This may include correction of DNA sequencing errors, normalisation of gene expression data, removal of low-quality data or handling of missing values.
- Statistical analysis: Once pre-processed, the data are subjected to statistical analysis to identify patterns, relationships or significant differences. This may involve the use of statistical tests, such as the t-test or analysis of variance (ANOVA), to compare sample groups or the calculation of correlations between different variables.
- Identification of significant features: In many cases, it is necessary to identify features or biological variables that are significantly associated with a given phenomenon or condition. This can be done using feature selection methods such as principal component analysis (PCA), logistic regression or decision tree.
- Interpretation of results: Once the analyses have been completed, it is important to interpret the results in light of the research questions or objectives of the experiment. This may require the use of biological databases, bioinformatics tools or consultation of the scientific literature to gain a better understanding of the biological implications of the results.
- Data visualisation: The use of graphs and visualisations is an important aspect of experimental and bioinformatics data analysis. Visualisations can help to identify patterns or trends in the data, making it easier to communicate results in a clear and understandable way.

This course explores topics closely related to one or more goals of the United Nations 2030 Agenda for Sustainable Development (SDGs).

## Obiettivi per lo sviluppo sostenibile

Codice	Descrizione
3	Good health and well-being