UNIVERSITÀ
DEGLI STUDI DI MILANO-BICOCCA

SYLLABUS DEL CORSO

Laboratorio di Biostatistica ed Epidemiologia

2627-3-E4102B092

Learning objectives

The aim of the course is to provide the methodological and operational tools necessary to set up, manage and analyze biomedical and epidemiological data, with particular attention to the transition from research protocol to analytical dataset, survival data analysis, introduction to diagnostic and prognostic models, data quality control and production of tables, descriptive figures and short statistical reports.

Knowledge and understanding
This course will provide knowledge and understanding of the structure of a research protocol and its translation into Case Report Form (CRF) and analytical dataset; principles of sample size and statistical power; methods for the analysis of survival data; introductory elements of diagnostic and prognostic models; principles of data quality control; criteria for the production of tables and descriptive figures and for the preparation of a short statistical report.

Ability to apply knowledge and understanding
At the end of the course, students will be able to: set up the essential elements of a dataset starting from a research protocol; recognize the essential elements related to the sample size and statistical power; analyze survival data using Kaplan-Meier curves, log-rank tests and Cox models; interpret survival curves, hazard ratios and assumption of proportionality of risks; understand the role of diagnostic and prognostic models in risk classification and prediction; carry out completeness, plausibility and consistency checks on the data; identify outliers, missing data, and inconsistencies; produce tables and descriptive figures consistent with the type of study and the nature of the variables; draw up a short statistical report.

Making judgements
Through practical exercises on biomedical and epidemiological data, students will develop critical skills in assessing data quality, choosing appropriate statistical methods and interpreting results, taking into account the study design, the structure of the dataset, the nature of the variables and the limitations of the analyses conducted.

Communication skills
Students will be guided in the production of tables, figures and short texts commenting on the results. The drafting

of a short statistical report will allow students to develop the ability to communicate clearly and with appropriate technical language the objectives, data, methods, results and limits of the analysis.

Learning skills
The course will provide students with conceptual and operational tools to independently continue the deepening of statistical methods applied to biomedical and epidemiological research, through the use of statistical software, exercise datasets, scripts and technical documentation.
The course allows the student to acquire applied skills in the use of statistics in the biostatistical and epidemiological work context.

Contents

1. From the research protocol to the analytical dataset
2. Survival Data Analysis
3. Introduction to diagnostic and prognostic models
4. Data quality control, tables and descriptive figures, statistical report

Detailed program

1. From the research protocol to the analytical dataset
Essential structure of a research protocol in the biomedical and epidemiological fields. Definition of the research question, study population, inclusion and exclusion criteria, exposure, study intervention or test, outcomes, and covariates. Sample size and statistical power. Operational differences between prospective and retrospective studies in data collection. Translation of the protocol into a CRF. General principles of organization of the analytical dataset.

2. Survival Data Analysis
Characteristics of time-to-event data. Source time, follow-up time, event, censorship, and truncation. Survival function, risk function and cumulative risk. Nonparametric estimation of survival by Kaplan-Meier method. Graphical representation and interpretation of survival curves. Comparison between groups using log-rank tests. Cox model with proportional risks, inclusion of covariates, interpretation of the hazard ratio and evaluation of the assumption of proportionality of hazards.

3. Introduction to diagnostic and prognostic models
Aims and areas of application of diagnostic and prognostic models in the biomedical and epidemiological fields. Difference between diagnosis of a present condition and prognosis of a future event. Definition of the target population, outcome, predictors and prediction time horizon. Construction, interpretation and evaluation of statistical models for risk classification and prediction. Introduction to the concepts of predictive performance, validation and applicability of models in clinical and epidemiological practice. Introduction to the use of survival data models in the estimation of individual risk at fixed times.

4. Data quality control, descriptive tables and statistical reporting
Completeness, plausibility and consistency checks on the data collected. Identify and manage outliers, missing data, and inconsistencies. Construction of the final analytical dataset. Production of descriptive tables consistent with the type of study and with the nature of the variables under analysis. Structure of a short statistical analysis report.

Prerequisites

Knowledge of descriptive and inferential statistics: probability distributions, estimation, confidence intervals, hypothesis testing and p-values. Knowledge of the contents of the Medical Statistics course is also required, in particular measures of frequency and association, main study designs and interpretation of the results of biomedical research.

Teaching methods

Lessons will take place in two ways.
Delivery methods: presentation of methodological and application concepts by the teacher, with introduction of the statistical tools necessary to deal with the different topics of the course.
Interactive mode: computer lab activities, guided exercises on real or simulated data, analysis using SAS statistical software, discussion of outputs and interpretation of results.
Indicatively, the ratio between the hours in delivery and interactive mode will be 1:2.

Assessment methods

There are no mid-term tests.
The exam will take place in a single day and will be divided into two sections:
• in the first section, students will have to answer in writing to closed and/or open questions on the topics of the course;
• in the second section, on the computer, one or more problems based on real or simulated data will be presented, to be solved using the SAS statistical software.

The first section is aimed at verifying the knowledge and understanding of the methodological concepts covered in the course, as well as the ability to correctly interpret them in the biomedical and epidemiological fields.
The second section is aimed at verifying the ability to apply statistical methods to real or simulated data, produce and interpret statistical outputs, carry out checks on the data, build tables and graphs and comment appropriately on the results obtained.
The evaluation will take into account the methodological correctness of the answers and analyses carried out, the appropriateness of the statistical procedures used, the ability to critically interpret the outputs, the clarity of exposition and the correct use of technical language.

The examination method is identical for attending and non-attending students.

Textbooks and Reading Materials

The teaching material will be made available on the e-learning platform of the course and will include slides, exercises, datasets, scripts, scientific articles and further in-depth materials.

Semester

II semester, III cycle

Teaching language

Italian.
Part of the teaching material (scientific articles, software documentation, references) will be in English.

Sustainable Development Goals

GOOD HEALTH AND WELL-BEING