

CORSO DI STUDIO: SCIENZE CHIMICHE (LAUREA MAGISTRALE)

ANNO ACCADEMICO: 2023-2024

**DENOMINAZIONE DELL'INSEGNAMENTO:
CHEMIOMETRIA / CHEMOMETRICS**

Principali informazioni sull'insegnamento	
Anno di corso	2023/2024
Periodo di erogazione	Secondo semestre (Marzo 2024 – Giugno 2024)
Crediti formativi universitari (CFU/ETCS):	6
SSD	Chimica Analitica – CHIM01
Lingua di erogazione	Italiano
Modalità di frequenza	Non obbligatoria

Docente	
Nome e cognome	Ilario Losito
Indirizzo mail	ilario.losito@uniba.it
Telefono	080-5442506
Sede	Dipartimento di Chimica, stanza n. 11
Sede virtuale	Classe Teams con codice a4q5nck
Ricevimento	Nella classe Teams su indicata, previa prenotazione via e-mail

Organizzazione della didattica			
Ore			
Totali	Didattica frontale	Pratica (laboratorio, campo, esercitazione, altro)	Studio individuale
150	32	30	88
CFU/ETCS			
6	4	2	

Obiettivi formativi	Conoscenza dei fondamenti teorici e delle applicazioni dei principali metodi statistici e chemiometrici di impiego in ambito chimico.
Prerequisiti	Nozioni di statistica descrittiva acquisite nel Corso di Laurea Triennale in Chimica (in particolare nell'insegnamento di Laboratorio di Chimica Analitica II).

<p>Metodi didattici</p>	<p>Lezioni frontali mediante presentazioni con software PowerPoint</p>
<p>Risultati di apprendimento previsti</p>	<ul style="list-style-type: none"> - Descrittore di Dublino 1: conoscenza e capacità di comprensione; <ul style="list-style-type: none"> ○ Conoscenza dei principali aspetti della statistica inferenziale in chimica analitica ○ Conoscenza dei fondamenti teorici dei principali metodi chemiometrici utilizzati in ambito scientifico. ○ Conoscenza degli aspetti salienti del controllo di qualità e del campionamento - Descrittore di Dublino 2: capacità di applicare conoscenza e comprensione; <ul style="list-style-type: none"> ○ Progettare, selezionare ed ottimizzare procedure ed esperimenti ○ Estrarre la massima informazione possibile sul sistema in esame attraverso l'analisi e l'interpretazione dei dati ○ Valutare problematiche inerenti il controllo qualità e il campionamento - Descrittore di Dublino 3: capacità critiche e di giudizio <ul style="list-style-type: none"> • <i>Autonomia di giudizio</i> <ul style="list-style-type: none"> ○ Capacità di affrontare su base chemiometrica l'analisi di un problema complesso. - Descrittore di Dublino 4: capacità di comunicare quanto si è appreso <ul style="list-style-type: none"> • <i>Abilità comunicative</i> <ul style="list-style-type: none"> ○ Esporre in modo chiaro l'impostazione e l'interpretazione dei risultati ottenuti dopo l'applicazione di un metodo chemiometrico - Descrittore di Dublino 5: capacità di proseguire lo studio in modo autonomo nel corso della vita <ul style="list-style-type: none"> • <i>Capacità di apprendere in modo autonomo</i> <ul style="list-style-type: none"> ○ Capacità di trasferire le conoscenze di base acquisite a nuove problematiche applicative anche a carattere interdisciplinare.

<p>Contenuti di insegnamento (Programma)</p>	<p>Cenni storici e concetti generali sulla chemiometria</p> <p>Statistica descrittiva e inferenziale</p> <p>Variabili random discrete e continue, funzioni densità di probabilità e di distribuzione cumulativa. Rappresentazioni grafiche di distribuzioni univariate, momenti di una distribuzione, misure di tendenza centrale e di dispersione, indici di forma: asimmetria e curtosi, entropia di una distribuzione, funzione generatrice dei momenti, distribuzioni di uso più frequente.</p> <p>Inferenza statistica: concetti generali. Approccio frequentista: stima puntuale di un parametro di popolazione, intervallo di fiducia. Approccio Bayesiano: teorema di Bayes, intervalli credibili. Confronto fra intervalli di fiducia ed intervalli credibili.</p> <p>Richiami sui test di ipotesi: errori di tipo I e II e potenza dei test. Doppio t-test ad una cosa (TOST) per la valutazione dell'equivalenza. Confronto fra t-test e TOST. Dati aberranti: generalità. Test Q di Dixon per uno o più dati aberranti. Test di Grubbs. Individuazione dei dati aberranti mediante i Box-and-Whiskers plot di Tukey. Deviazione assoluta dalla mediana (MAD).</p> <p>Test per il confronto fra più varianze: F_{\max} test di Hartley; test di Bartlett; S test di Levene.</p> <p>Analisi della varianza (ANOVA) e confronto multiplo fra medie</p> <p>ANOVA ad una via: concetti generali. Ripartizione della variabilità totale. Somme dei quadrati e Quadrati Medi. Valori attesi per i Quadrati Medi nell'ANOVA ad una via. Fattori fissi e fattori random nell'ANOVA ad una via.</p> <p>Confronto multiplo fra medie: concetti generali. Test di Fisher e di Fisher-Hayter sulla minima differenza significativa, test di Tukey, Tukey-Kramer, Bonferroni, multiple range test di Duncan, test di Dunnett per confronti con un controllo. Confronto multiplo fra medie mediante metodi non parametrici: test di Kruskal-Wallis e test di Dunn.</p> <p>Metodi di regressione</p> <p>Regressione lineare semplice. Varianze e distribuzioni degli stimatori di intercetta e pendenza. ANOVA relativa alla regressione lineare. Coefficienti di determinazione e di correlazione. Intervalli di fiducia e di predizione per i responsi sottoposti a regressione lineare. Mancanza di adattamento nella regressione semplice lineare. Calibrazione e regressione inversa. Intervalli di Wald e di inversione. Disegno sperimentale per la calibrazione.</p> <p>Regressione lineare univariata di ordine m: equazioni matriciali, stima della varianza, speranza matematica e varianza del vettore dei coefficienti, intervalli di fiducia.</p> <p>Regressione lineare multipla: stima della varianza, speranza matematica e varianza del vettore dei coefficienti, intervalli di fiducia, multicollinearità, coefficiente di determinazione. Criteri per la scelta dei regressori nella regressione lineare multipla. Applicazione della regressione lineare multipla all'analisi multicomponente. Deriva e validazione di un modello. Tecniche di validazione. Validazione dei metodi di regressione. Parametri che misurano la capacità predittiva di un modello. Metodi diagnostici per la regressione: residui e valori-leva. Dati aberranti e dati influenti nella regressione lineare semplice.</p> <p>Disegno sperimentale ed ottimizzazione</p> <p>Disegno sperimentale: generalità. Problemi con l'approccio "una variabile per</p>
---	--

	<p>volta". Stadi per il disegno di un esperimento. Modelli empirici. Disegni di screening: disegni fattoriali completi e frazionati; disegno di Plackett-Burman.</p> <p>Ottimizzazione di processo con il disegno sperimentale. Operazione evolutiva di tipo Box. Ottimizzazione mediante semplice. Metodo dell'ascesa più ripida. Metodi della superficie di risposta. Disegni compositi centrali. Disegni di Box-Behnken. Campi di applicazione del disegno sperimentale.</p> <p>Disegno miscela: considerazioni generali. Tipi di disegno miscela: semplice-reticolo, semplice-centroide, a vertici estremi. Modelli di primo, secondo e terzo grado. Uso del disegno sperimentale nella scienza delle separazioni: disegni di screening, approccio del semplice, metodi basati sulla superficie di risposta. Applicazione del disegno miscela all'ottimizzazione delle fasi mobili in HPLC.</p> <p>Metodi multivariati</p> <p>Concetti generali sui metodi multivariati. Matrici di covarianza e di correlazione. Analisi delle Componenti Principali (PCA): aspetti generali, determinazione degli autovalori e degli autovettori, proiezione dei dati, interpretazione geometrica delle componenti principali. Stadi di un'elaborazione PCA. Risultati di un'elaborazione PCA: score, loadings e scree plot, cerchio della correlazione. Esempi di elaborazioni PCA.</p> <p>Calibrazione inversa nell'analisi multi-componente: uso della Regressione Lineare Multipla (MLR). Regressione delle Componenti Principali (PCR). Regressione con i Minimi Quadrati Parziali (PLS). Confronto fra gli approcci MLR, PCR e PLS.</p> <p>Analisi dei Cluster (CA): concetti generali. Analisi dei Cluster Gerarchica (HCA): approcci agglomerativi e divisivi, tipi di distanza impiegati più comunemente, matrici delle distanze e delle somiglianze. Algoritmi di raggruppamento nella HCA: legame singolo, completo, medio e fra centroidi, metodo di Ward. Metodi di raggruppamento non gerarchici: algoritmo delle k medie.</p> <p>Metodi di classificazione: concetti generali, matrice di confusione, probabilità a priori, sensibilità e specificità. Metodo dei k vicini più prossimi (K-NN). Analisi discriminante: concetti generali. Analisi Discriminante Lineare (LDA): proiezione dei dati su un nuovo asse, funzione discriminante lineare di Fisher e sua massimizzazione. Classificazione di un nuovo oggetto mediante LDA. LDA con più di due classi. Limitazione della LDA e confronto con la PCA.</p> <p>Cenni su controllo di qualità e campionamento</p> <p>Qualità delle misure analitiche, metodi per il controllo di qualità. Carte di controllo Shewhart per i valori medi. Lunghezza media di una corsa (ARL). Carte di controllo CUSUM. Test di competenza. Prove collaborative inter-laboratorio. Valutazione dell'incertezza complessiva su una misura.</p> <p>Campionamento: concetti generali. Aspetti chiave di un piano di campionamento: luogo/tempo del campionamento dalla popolazione target, tipo di campioni, minima quantità di campione per analisi, numero di campioni da analizzare, minimizzazione della varianza totale associata all'analisi. Cenni sul campionamento per l'accettazione basato su variabili.</p>
<p>Testi di riferimento</p>	<p>M. Forina, Fondamenta per la Chimica Analitica, e-book, ISBN 9788890406461.</p> <p>L. Saliani, Manuale di statistica per la ricerca e la professione, 2005 (disponibile online).</p> <p>R. Todeschini, Introduzione alla Chemiometria, EdiSES, 1998.</p> <p>D.C. Montgomery, Design and analysis of experiments, 9a edizione, John Wiley & Sons Inc., 2017.</p> <p>J.N. Miller, J.C. Miller e R.D. Miller, Statistics and chemometrics for analytical chemistry, 7a edizione, Pearson Prentice Hall, 2017.</p> <p>M. Otto, Chemometrics - Statistics and Computer Application in Analytical Chemistry, 3a edizione, Wiley-VCH, 2017.</p> <p>D. Harvey, Analytical Chemistry 2.0, eText, Chapter 7, 2008</p> <p>S.L.R. Ellison, V.J. Barwick, T.J. Duguid Farrant, Practical Statistics for the</p>

	Analytical Scientist, 2a edizione, RSC Publishing, 2009.
Note ai testi di riferimento	Il docente chiarisce a lezione quali parti dei testi suddetti possono fungere da supporto nello studio degli argomenti del corso.
Materiali didattici	Il materiale didattico usato a lezione è disponibile per il download, in forma di file PDF, da apposito sito web, comunicato agli studenti all'inizio del corso.
Valutazione	
Modalità di verifica dell'apprendimento	Prova orale con una domanda per ciascuna macroarea del programma di insegnamento

<p>Criteri di valutazione</p>	<ul style="list-style-type: none"> • <i>Conoscenza e capacità di comprensione:</i> <ul style="list-style-type: none"> o Conoscenza completa e ordinata dell'argomento inerente la domanda posta dal docente su ciascuna macro-area del programma • <i>Conoscenza e capacità di comprensione applicate:</i> <ul style="list-style-type: none"> o Risposta a domande di natura applicativa indotte dall'esposizione iniziale dell'argomento • <i>Autonomia di giudizio:</i> <ul style="list-style-type: none"> o Capacità di confrontare autonomamente diverse tecniche statistiche in termini di pro e contro • <i>Abilità comunicative:</i> <ul style="list-style-type: none"> o Esposizione in modo ordinato e consequenziale dell'argomento oggetto di domanda • <i>Capacità di apprendere:</i> <ul style="list-style-type: none"> o Enunciazione e discussione, almeno in forma schematica, delle principali equazioni relative ai metodi statistici oggetto dell'insegnamento.
<p>Criteri di misurazione dell'apprendimento e di attribuzione del voto finale</p>	<p>Il voto finale, espresso in trentesimi, si baserà sulla valutazione della qualità delle risposte fornite alle domande sulle diverse macroaree del programma, tutte aventi lo stesso peso nella valutazione. La lode viene assegnata quando le risposte a tutte le domande risultano ineccepibili.</p>
<p>Altro</p>	

COURSE OF STUDY: CHEMICAL SCIENCES (MASTER'S DEGREE)
ACADEMIC YEAR: 2023/2024
ACADEMIC SUBJECT: CHEMOMETRICS

General information	
Year of the course	2023/2024
Academic calendar (starting and ending date)	Second semester (March 2023 – June 2023)
Credits (CFU/ETCS):	6
SSD	Analytical Chemistry – CHIM01
Language	Italian
Mode of attendance	Not mandatory

Professor/ Lecturer	
Name and Surname	Ilario Losito
E-mail	ilario.losito@uniba.it (illosdid@hotmail.com)
Telephone	080-5442506
Department and address	Chemistry, office n. 11
Virtual room	Teams class, code a4q5nck
Office Hours	In the above-mentioned Teams class, via e-mail booking

Work schedule			
Hours			
Total	Lectures	Hands-on (laboratory, workshops, working groups, seminars, field trips)	Out-of-class study hours/ Self-study hours
150	32	30	88
CFU/ETCS			
6	4	2	

Learning Objectives	Knowledge of theoretical fundamentals and applications of main chemometric methods employed in chemical contexts.
Course prerequisites	Notions of descriptive statistics gained during the Three-Year Degree Course in Chemistry (specifically, in the Laboratory of Analytical Chemistry II course).

Teaching strategies	Frontal lessons based on PowerPoint presentations
Expected learning outcomes in terms of	
Knowledge and understanding on:	<ul style="list-style-type: none"> ○ Knowledge of the main aspects of inferential statistics in analytical chemistry ○ Knowledge of theoretical fundamentals of the main chemometrics methods employed in the scientific context ○ Knowledge of main aspects of quality control and sampling
Applying knowledge and understanding on:	<ul style="list-style-type: none"> ○ Plan, select and optimize procedures and experiments ○ Draw the maximum amount of information on the system under study through data analysis and interpretation ○ Evaluate problems related to quality control and sampling
Soft skills	<ul style="list-style-type: none"> ● <i>Making informed judgments and choices</i> <ul style="list-style-type: none"> ○ Ability to deal with the analysis of a complex problem through a chemometric approach

- | | |
|--|---|
| | <ul style="list-style-type: none">• <i>Communicating knowledge and understanding</i><ul style="list-style-type: none">○ Express clearly the planning and the interpretation of results obtained after the application of a chemometric method• <i>Capacities to continue learning</i><p>Be able to transfer the acquired knowledges to new applicative problems, even with an interdisciplinary character.</p> |
|--|---|

Syllabus	
Content knowledge	<p>Historical background and general concepts of chemometrics</p> <p>Descriptive and inferential statistics</p> <p>Discreet and continuous random variables, probability density and cumulative distribution functions. Graphical representations of univariate distributions, moments of a distribution, central tendency and dispersion indicators, shape indexes: asymmetry and curtosis, entropy of a distribution, moment generating function, more frequently used distributions.</p> <p>Statistical inference: general concepts. Frequentist approach: point estimate of a population parameter, confidence interval. Bayesian approach: Bayes theorem, credible intervals. Comparison between confidence and credible intervals.</p> <p>References to hypothesis testing: type I and II errors, power of tests. Two one-sided t-test (TOST) for the evaluation of equivalence. Comparison between test and TOST.</p> <p>Outliers: general concepts. Dixon Q-test for one or more outliers. Grubbs test. Recognition of outliers by Tukey Box-and Whiskers plot. Median Absolute Deviation (MAD).</p> <p>Test for the comparison of several variances: Hartley F_{\max} test, Bartlett test, Levene S test.</p> <p>Analysis of Variance (ANOVA) and multiple comparison between means</p> <p>One-way ANOVA: general concepts. Partition of total variability. Sums of Squares and Mean Squares. Expected values for Mean Squares in one-way ANOVA. Fixed and random factors in one-way ANOVA.</p> <p>Multiple comparison between means: general concepts. Fisher and Fisher-Hayter least significant difference tests, Tukey, Tukey-Kramer, Bonferroni tests, Duncan multiple range test, Dunnett test for comparisons with a control. Multiple comparison between means based on non parametric methods: Kruskal-Wallis test and Dunn test.</p> <p>Regression methods</p> <p>Simple linear regression, Variances and distributions of intercept and slope estimators. ANOVA for linear regression. Determination and correlation coefficients. Confidence and regression intervals for responses subjected to linear regression. Lack of fit in simple linear regression. Calibration and inverse regression. Wald and inversion intervals. Experimental design for calibration.</p> <p>Univariate linear regression of order m: matrix equations, estimate of variance, expectation and variance of the coefficients vector, confidence intervals.</p> <p>Multiple linear regression: estimate of variance, expectation and variance of the coefficients vector, confidence intervals, multicollinearity, determination coefficient. Criteria for the choice of regressors in multiple linear regression. Application of multiple linear regression to multicomponent analysis.</p> <p>Bias and validation of a model. Validation techniques. Validation of regression methods. Parameters measuring the predictive ability of a model. Diagnostic methods for regression: residual and leverage values. Outlying and influential data in simple linear regression.</p> <p>Experimental design and optimization</p> <p>Experimental design: general concepts. Problems with the "one variable at a time" approach. Steps for the design of an experiment. Empirical models. Screening</p>

	<p>designs: complete and fractional factorial design; Plackett-Burman design. Process optimization with experimental design. Box evolutionary operation. Simplex optimization. Steepest ascent method. Response surface methods. Central composite designs. Box-Behnken designs. Application fields for experimental design.</p> <p>Mixture design: general considerations. Types of mixture design: simplex-lattice, simplex-centroid, extreme vertices. First, second, and third-degree models. Use of experimental design in separation science: screening designs, simplex approach, surface response methods. Application of mixture design to the optimization of mobile phases in HPLC.</p> <p>Multivariate methods</p> <p>General concepts on multivariate methods. Covariance and correlation matrices. Principal Components Analysis (PCA): general aspects, determination of eigenvalues and eigenvectors, data projection, geometric interpretation of principal components. Steps of a PCA elaboration. Results of a PCA elaboration: score, loadings and scree plots, correlation circle. Examples of PCA elaborations. Inverse calibration in multi-component analysis: use of Multiple Linear Regression (MLR). Principal Components Regression (PCR). Partial Least Squares (PLS) Regression. Comparison between MLR, PCR and PLS approaches.</p> <p>Cluster Analysis (CA): general concepts. Hierarchical Cluster Analysis (HCA): agglomerative and divisive approaches, most commonly adopted types of distance, matrices of distances and similarities. Grouping algorithms in HCA: single, complete, average and centroid linkage, Ward method. Non-hierarchical grouping methods: k-means algorithm.</p> <p>Classification methods: general concepts, confusion matrix, <i>a priori</i> probability, sensitivity and specificity. K-nearest neighbours (K-NN) method. Discriminant analysis: general concepts. Linear Discriminant Analysis (LDA): data projection on a new axis, Fisher linear discriminating function and its maximization. Classification of a new object by LDA. LDA with more than two classes. Limitations of LDA and comparison with PCA.</p> <p>Hints on quality control and sampling</p> <p>Quality of analytical measurements, methods for quality control. Shewhart control charts for mean values. Average run length (ARL). CUSUM control charts. Proficiency testing. Inter-laboratory collaborative trials. Evaluation of the total uncertainty of a measurement.</p> <p>Sampling: general concepts. Key aspects of a sampling plan: place/time of sampling of target population, types of samples, minimal amount of sample for the analysis, number of samples to analyse, minimization of total variance associated to the analysis. Hints on acceptance sampling based on variables.</p>
<p>Texts and readings</p>	<p>M. Forina, <i>Fondamenta per la Chimica Analitica</i>, e-book, ISBN 9788890406461.</p> <p>L. Saliani, <i>Manuale di statistica per la ricerca e la professione</i>, 2005 (disponibile online).</p> <p>R. Todeschini, <i>Introduzione alla Chemiometria</i>, EdiSES, 1998.</p> <p>D.C. Montgomery, <i>Design and analysis of experiments</i>, 9a edizione, John Wiley & Sons Inc., 2017.</p> <p>J.N. Miller, J.C. Miller e R.D. Miller, <i>Statistics and chemometrics for analytical chemistry</i>, 7a edizione, Pearson Prentice Hall, 2017.</p> <p>M. Otto, <i>Chemometrics - Statistics and Computer Application in Analytical Chemistry</i>, 3a edizione, Wiley-VCH, 2017.</p> <p>D. Harvey, <i>Analytical Chemistry 2.0</i>, eText, Chapter 7, 2008</p> <p>S.L.R. Ellison, V.J. Barwick, T.J. Duguid Farrant, <i>Practical Statistics for the Analytical Scientist</i>, 2a edizione, RSC Publishing, 2009.</p>
<p>Notes, additional materials</p>	<p>During classes the teacher clarifies which parts of the above mentioned books can be a support for studying the course topics.</p>

Repository	The website from which didactic materials can be downloaded, as PDF files, will be indicated to students at the beginning of the course
Assessment	
Assessment methods	Oral exam, with a question for each macro-area of the course program
Assessment criteria	<ul style="list-style-type: none"> • <i>Knowledge and understanding</i> <ul style="list-style-type: none"> ○ Complete and ordered knowledge of the topic concerning the question posed by the teacher on each macro-area of the course program • <i>Applying knowledge and understanding</i> <ul style="list-style-type: none"> ○ Answer to questions on applicative aspects arising from the initial presentation of a topic • <i>Autonomy of judgment</i> <ul style="list-style-type: none"> ○ Ability to compare independently different statistical techniques in terms of pros and cons • <i>Communication skills</i> <ul style="list-style-type: none"> ○ Express the topic object of a question in an ordered and consequential way • <i>Capacities to continue learning</i> <ul style="list-style-type: none"> ○ Enunciation and discussion, at least in a schematic form, of the main equations related to statistic methods object of the course.
Final exam and grading criteria	The final mark, expressed in thirtieths, will be based on the evaluation of quality of answers given to questions on different macro-areas of the program, all having the same weight in the evaluation. Honors are assigned when the answers to all questions are irreproachable.
Further information	