

Elaborazione dati e visualizzazione grafica dell'informazione in campo ambientale e sanitario

Tecniche statistiche, geostatistiche e chemiometriche per la rappresentazione grafica dell'informazione contenuta in grandi data set ai fini del controllo ambientale e sanitario

Destinatari: studenti del III anno del CdL in *Tecniche della Prevenzione nell'ambiente e nei luoghi di lavoro* dell'Università di Trieste¹

Durata: 15-20 ore; A.A. 2021-2022 I semestre (ottobre-dicembre)

Tutor: **Andrea Mistaro** (ARPA FVG – Sede di Trieste)

Programma indicativo:

1. Dalla progettazione del campionamento al *data analysis*: la “ruota” di Vandeginste.
2. Estrazione dati: il *data mining* e i problemi connessi; formattazione del database.
3. Chemiometria: estrazione dell'informazione rilevante dai dati:
 - *Exploratory Data Analysis* (EDA) vs *Confirmatory Statistics* (statistica descrittiva)
4. Reporting techniques: visualizzazione grafica dei dati;
 - Rappresentazione grafica e tabellare a confronto;
 - *Graphical perception* (cenni teorici);
 - Integrazione dei grafici nel testo; infografica (cenni).
5. Analisi univariata: cenni di statistica descrittiva e relative tecniche di *reporting* grafico:
 - La distribuzione di frequenza e le sue tecniche di rappresentazione (istogrammi; curve di distribuzione; *Box & Whisker plot*);
 - Serie temporali (introduzione e peculiarità).
6. Analisi bivariata: analisi delle correlazioni e relative tecniche di *reporting* grafico:
 - Coefficienti di correlazione; *correlation plots*; correlazione vs regressione; regressione lineare (retta dei minimi quadrati e significato “fisico” dei parametri di regressione); cenni sui rapporti diagnostici tra inquinanti.
7. Analisi multivariata:
 - *Principal component analysis* (PCA);
 - Tecniche di classificazione (LDA, SIMCA);
 - Cenni di *cluster analysis*, reti neurali e *self-organizing maps*.
8. Data Maps: cenni di geostatistica: l'interpolazione spaziale dei dati e relative tecniche di *reporting* grafico.
9. Time Series: approfondimento sull'analisi di serie temporali:
 - analisi preliminari: indici di continuità e completezza della serie storica;
 - carte di controllo (*control charts*); condizioni per l'impiego della regressione lineare; *smoothing techniques* (LoWeSS); autocorrelogrammi; cross-correlogrammi.
10. Cenni di modellistica previsionale.

¹ Evoluzione del precedente seminario/tirocinio “*Estrazione e visualizzazione dell'informazione ambientale contenuta in grandi data set con tecniche statistiche, geostatistiche e chemiometriche a fini del controllo ambientale*” per gli studenti del CdL in *Tecniche della Prevenzione nell'ambiente e nei luoghi di lavoro* dell'Università di Trieste (12/1/2011, 3 ore; 30/1/2013, 3,5 ore; 12/1/15, 4 ore; 3/2/2017, 4 ore; A.A. 2017-18, 12 ore)

Programma delle esercitazioni al PC

*lo stesso giorno della rispettiva parte teorica,
gli studenti saranno invitati a svolgere delle esercitazioni al PC
(in coppie o terne di studenti)
su alcuni degli argomenti trattati, con la supervisione del docente*

- 5a. Esercitazioni di analisi statistica descrittiva classica (univariata) e relativo *reporting* grafico:
 - distribuzioni di frequenza: istogrammi; curve di distribuzione; *Box & Whisker plot*; *normal probability plot*.
- 6. Esercitazioni di analisi delle correlazioni (analisi bivariata) e relativo *reporting* grafico:
 - plot bivariati; coefficienti di correlazione; *correlation plots*; regressione lineare;
 - Analisi del “quartetto di Anscombe”.
- 7a. Analisi multivariata
 - *Principal component analysis* (PCA).
- 9. Esercitazioni di analisi e *reporting* grafico delle serie temporali²:
 - *Control charts* (carta di controllo della media);
 - LoWeSS;
 - Autocorrelogramma.

Project-working finale

*da svolgere autonomamente e senza supervisione del docente,
divisi in 5 gruppi di lavoro da 3 studenti ciascuno;
con elaborato finale da consegnare per correzione e verifica dell'apprendimento*

- Libera analisi, con le tecniche ritenute più idonee, di un caso ambientale reale, con produzione di un breve report (max 6 pag. di testo con grafici + max 12 slide .ppt);
- Da parte degli studenti: successiva esposizione in aula (con slide) dei risultati;
- Da parte del docente: commento in aula degli elaborati e delle esposizioni.

Risorse necessarie:

- Aula con PC / MS Office Powerpoint e videoproiettore per il docente.
- Aula informatica per le esercitazioni (un PC per studente o – per alcune esercitazioni – uno per “trio”), con installato il pacchetto Office.
- Possibilità di installare s/w specifici gratuiti (Past.exe) quantomeno sui PC degli studenti.

² Ogni gruppo di studenti procurerà delle serie temporali di misurazioni a scelta; seguono alcuni esempi di serie temporali idonee come dataset di esercitazione:

- serie “periodiche” (es: serie storiche di temperatura esterna con variazioni stagionali o diurne...)
- serie non periodiche (es: rilevamento e analisi dell'andamento della temperatura dell'acqua di rete; della temperatura corporea; andamento della concentrazione di un inquinante ambientale – es: PM10 in aria ambiente).