

OEC019

Analisi dati esplorativa (con R o Python)

Durata: 1 gg

Descrizione

L'esame sistematico dei dati a disposizione mediante tecniche statistiche, per scoprire trend, pattern e relazioni, e presentarli efficacemente ai Manager con R. L'analisi esplorativa dei dati aziendali disponibili (alias Data Discovery) è in genere il passo preliminare alla modellazione, ed è dunque fondamentale nell'Analitica avanzata e nella Data Science in generale.

3 passi classici (per partire): verifica della distribuzione dei valori, verifica delle relazioni tra risposta e predittori, verifica delle relazioni tra predittori. Taglio pratico con il software R od, in alternativa su richiesta, con Excel 2013+.

Prerequisiti

Il corso introduttivo su R.

Contenuti

- Tipi di variabili (numeriche e categoriche).
 - Misure di sintesi dei dati.
 - Istogrammi e tabelle di frequenza.
 - Misure centrali dei dati.
 - Misure di dispersione dei dati.
 - Rappresentazione grafica dei dati.
 - Campione vs popolazione.
 - Distribuzioni di probabilità (le principali per le aziende).
 - Correlazione e covarianza.
 - Regressione lineare semplice.
 - Assunzioni ed analisi dei residui.
 - Rilevamento, analisi e trattamento degli outlier (valori estremi) e dei valori mancanti.
 - Regressione linearizzabile.
 - Regressione non lineare.
 - Regressione multipla e multivariata (cenni).
 - Regressione stocastica (cenni).
 - Analisi della varianza (cenni).
 - Intervalli di confidenza e di previsione (per correlazione e regressione). Riduzione variabili / PCA. Analisi dati esplorativa per dati non-strutturati (cenni). Test di ipotesi. Grafici e plot (scatterplot, boxplot, density plot). Visualizzazione grafica dei dati (avanzata): cenni.
- NB. Gli esempi relativi ai suddetti problemi sono in R. Quali tool usare per i suddetti problemi, oltre ad R: rapida rassegna (in primis Excel e Microsoft in generale) e confronto.