

A.A. 2020/2021

Case study – Regressione lineare

Corso di Statistica e Modelli Stocastici (SMS1)

Corso di Statistica e Topografia (Modulo di Statistica)

Prof. Alessandro Fassò e Dott. Paolo Maranzano

Ad ogni gruppo è stato assegnato un dataset GXX.xlsx o GXX.mat che contiene i dati settimanali di qualità dell'aria (inquinanti) e meteorologia misurati dal 2017 a fine settembre 2020 in due stazioni della Lombardia. La prima stazione è unica per ogni gruppo, mentre la seconda è comune a tutti i gruppi e riguarda i dati osservati a Bergamo in via Meucci (stazione rilevamento ARPA). Le variabili contenute sono:

- **Data**: data in cui sono stati rilevate le misurazioni. Formato giorno/mese/anno;
- **ARPA_AQ_cod_staz_tG1**: codice ARPA della stazione unica per ogni gruppo
- **ARPA_AQ_nome_staz_tG1**: nome (comune) della stazione ARPA unica per ogni gruppo
- **PM10_tG1**: concentrazione di PM_{10} in $\mu g/m^3$ osservati nella stazione unica;
- **NOX_tG1**: concentrazione di NO_x in $\mu g/m^3$ osservati nella stazione unica;
- **NO2_tG1**: concentrazione di NO_2 in $\mu g/m^3$ osservati nella stazione unica;
- **O3_tG1**: concentrazione di O_3 in $\mu g/m^3$ osservati nella stazione unica;
- **Temperatura_tG1**: temperatura media in Celsius nella stazione unica;
- **Pioggia_Cumulata_tG1**: quantità di pioggia totale in mm precipitata nella stazione unica;
- **Umidità_relativa_tG1**: umidità relativa in % osservata nella stazione unica.
- **ARPA_AQ_cod_staz_BG**: codice ARPA della stazione di Bergamo
- **ARPA_AQ_nome_staz_BG**: nome (comune) della stazione di Bergamo
- **PM10_BG**: concentrazione di PM_{10} in $\mu g/m^3$ osservati nella stazione di Bergamo;
- **NOX_BG**: concentrazione di NO_x in $\mu g/m^3$ osservati nella stazione di Bergamo;
- **NO2_BG**: concentrazione di NO_2 in $\mu g/m^3$ osservati nella stazione di Bergamo;
- **O3_BG**: concentrazione di O_3 in $\mu g/m^3$ osservati nella stazione di Bergamo;
- **Temperatura_BG**: temperatura media in Celsius nella stazione di Bergamo;
- **Pioggia_Cumulata_BG**: quantità di pioggia totale in mm precipitata nella stazione di Bergamo
- **Umidità_relativa_BG**: umidità relativa in % osservata nella stazione di Bergamo.

È stata messa a disposizione anche una ulteriore tabella (comune a tutti i gruppi) chiamata Anagrafica_stazioni.txt in cui sono elencate alcune informazioni sulla stazione di riferimento: comune, provincia, altitudine, coordinate geografiche e tipologia di stazione.

Consegna: Utilizzare tecniche di regressione lineare multipla in Matlab o Excel per investigare le concentrazioni osservate presso la stazione unica di ogni gruppo di uno tra gli inquinanti (NO₂, NO_x, PM₁₀, O₃) disponibili (variabile risposta) in funzione delle altre variabili a disposizione (Meteorologia oppure altri inquinanti).

Identificare il modello più adatto a descrivere le concentrazioni in base alle vostre conoscenze sui criteri di valutazione e confronto dei modelli (significatività, adattamento, trasformazioni dei dati, analisi dei residui).

Eseguire le medesime analisi sulla stazione di Bergamo e confrontare i risultati ottenuti tra le due stazioni. Alcune domande di interesse sono: le covariate dei due modelli sono uguali o differiscono? i coefficienti hanno gli stessi segni, valori simili e sono statisticamente significativi? I modelli hanno le stesse performances di adattamento?

Discutere le scelte effettuate e risultati ottenuti.

Output da consegnare:

- un report finale (in formato .pdf) con breve descrizione del lavoro svolto (max due pagine + eventuali figure) e i risultati principali e i vostri commenti (max due pagine + eventuali figure). Il file deve essere rinominato usando la seguente sintassi: GXX_YYYY_AssReg_Report.pdf, dove XX indica il numero del gruppo (es. 01, 02, ...) e YYYY indica il nome del gruppo (es. Andorra, Alaska, ...). Nella prima pagina del report devono essere indicati i nomi, cognomi e le matricole di tutti i componenti del gruppo, oltre che il nome del gruppo;
- un file .m/.xls con i calcoli effettuati in Matlab o Excel. Il file deve essere rinominato usando la seguente sintassi: GXX_YYYY_AssReg_calcoli.xlsx (Edili) oppure GXX_YYYY_AssReg_calcoli.m (Informatici), dove XX indica il numero del gruppo (es. 01, 02, ...) e YYYY indica il nome del gruppo (es. Andorra, Alaska, ...);

Modalità di consegna e scadenza: un rappresentante per ogni gruppo deve caricare nella repository di gruppo un file .zip o .rar denominato GXX_YYYY_AssReg.zip in cui sono inseriti il report pdf e i calcoli (.m o .excel) entro Domenica 10 gennaio 2021, ore 21.00

Discussione via Teams: la discussione da remoto (Teams) si svolgerà nella settimana dal lunedì 18 gennaio 2021 al venerdì 22 gennaio 2021.

Un rappresentante per ogni gruppo che intende presentare il proprio lavoro deve scrivere un messaggio Ilias a Paolo Maranzano indicando i nomi e la matricola dei membri e caricare la presentazione entro domenica 17 gennaio 2021 (ore 21.00). La presentazione può essere creata in formato .pdf o .ppt e deve essere rinominata usando la seguente sintassi: GXX_YYYY_AssReg_Presentazione.pdf. Oltre alla presentazione dei risultati, durante la discussione i docenti chiederanno ai membri del gruppo di rispondere ad alcune domande teoriche, valide per il voto orale, inerenti ai metodi utilizzati nel casi study.

Contatto: paolo.maranzano@guest.unibg.it

È preferibile utilizzare Ilias per le comunicazioni di gruppo: create un post nella cartella del gruppo in modo che tutti i componenti vedano i commenti. Nel caso vogliate comunicare tramite mail, mettete in cc tutti gli altri membri del gruppo.

Approfondimento sul tema inquinamento e meteo: sito di ARPA Lombardia

<https://www.arpalombardia.it/Pages/Aria/qualita-aria.aspx>

<https://www.arpalombardia.it/Pages/Meteorologia/Previsioni-e-Bollettini.aspx#/topPagina>