

Università degli Studi di Padova
Corso di Laurea Triennale in Statistica e Tecnologie Informatiche



ANALISI DI SERIE TEMPORALI DELLE CONCENTRAZIONI POLLINICHE

Relatore: Prof. (Dott.) Guido Masarotto
Dipartimento di Scienze Statistiche

Laureando: Gianluca Pollon
Matricola: 619473-STI

Anno Accademico 2011/2012

INDICE:

| | |
|---|-----------|
| I. Introduzione | 1 |
| I.1 Obiettivi | 1 |
| I.2 Osservazioni | 1 |
| 1. Il Polline | 2 |
| 1.1 Aerobiologia | 2 |
| 2. I Dati | 3 |
| 2.1 Somma Termica | 4 |
| 3. La strumentazione | 5 |
| 4. Metodologia ed Elaborazioni | 6 |
| 5. Scheda tecnica della Graminacea | 8 |
| 6. Analisi del polline della Graminacea | 9 |
| 7. Studio delle Somma Termica rispetto le Zone | 13 |
| 7.1 Preparazione del Data-Set | 14 |
| 7.2 Analisi | 16 |
| 7.3 Modellazione | 21 |
| 9. Altre attività svolte durante lo stage | 23 |
| 9.1 Open Data-Veneto Big Data | 23 |
| 9.2 Noise-Tube | 25 |
| 9.3 Corso di PHP | 26 |
| 12. Conclusioni e Ringraziamenti | 27 |
| 13 Sitologia e Bibliografia | 28 |

I. Introduzione

Questa relazione finale descrive il lavoro eseguito durante l'esperienza dello stage, svoltosi dal 15 Aprile al 15 Luglio; tre mesi con un totale di 350 ore di presenza presso l'ente pubblico Arpa Veneto con residenza in via Cairoli 4D, Padova.

Inizialmente il progetto è stato svolto singolarmente mentre nei due mesi successivi si è unito un compagno di facoltà, Simone Rossato Breda, per collaborare e proseguire lo studio.

Questo scritto riporta inoltre le attività effettuate durante lo stage non inerenti con l'obiettivo, ma comunque di livello formativo.

I.1 Obiettivi

- Effettuare un'analisi statistica descrittiva delle serie temporali delle concentrazioni di polline rilevate a partire dal 2001 fino ai giorni nostri in più stazioni venete e riguardanti diverse specie.
- Relazione del lavoro svolto concentrandosi su una determinata specie di polline: le graminacee, appartenenti alla famiglia delle Graminaceae.
- Confronto tra le serie storiche del polline e delle trasformazioni dei dati climatici con la finalità di evidenziarne correlazioni e identificare sistemi di previsione delle emissioni di polline in funzione dei dati meteorologici.

I.2 Osservazioni

Si tratta di un progetto molto oneroso in termini temporali, infatti oltre alle lunghe attese per ricevere i dati dall'azienda (non si disponeva dei diritti per interagire con le basi di dati), lo stesso processo di previsione richiede molto tempo per effettuare tentativi di modellazione. Ciò è dovuto al fatto che si ha a che fare con la natura, un'entità difficile da prevedere perché influenzata da molti aspetti, dunque si procede per tentativi, trovando le variabili che meglio permettono al modello di spiegare i dati.

Un esempio simile è riportato dall'Arpa Emilia che ha raggiunto l'obiettivo, ottenendo il modello predittivo in 8 anni di ricerche, utilizzando una rete neurale collegando più variabili. Questo esito, però, risulta molto correlato alla morfologia del territorio, pertanto non può essere riutilizzato nel territorio veneto, come difatti alcune prove hanno dimostrato.

1. Il Polline

Il polline è una particella biologica microscopica, grazie alla quale una notevole quantità di piante provvede alla propria riproduzione.

I pollini capaci di causare allergia sono detti allergenici e sono prodotti, nella maggioranza dei casi, da piante anemofile le quali affidano il loro polline al vento, perché possa raggiungere i fiori e le piante lontane anche decine di chilometri. Viceversa le piante entomofile si avvalgono di un sistema di trasporto “vivente”: gli insetti pronubi come l’ape o le formiche. Questo polline non contiene allergeni dunque non vengono eseguite analisi e monitoraggio, anche perché risulterebbero difficili.

Le piante anemofile producono particelle allergeniche aventi in genere un diametro inferiore ai 40-50 millesimi di millimetro quindi invisibili ad occhio nudo. Alcuni sono relativamente pesanti e possono essere ritrovati in un raggio di poche decine o centinaia di metri dalle piante da cui sono prodotti, altri, trasportati dal vento, possono raggiungere anche distanze di 10 km. I pollini compaiono in periodi ed in quantità differenti secondo la famiglia e la specie cui appartengono; la loro presenza dipende inoltre dalle caratteristiche climatiche e meteorologiche delle diverse località. La variabilità della data d’inizio della pollinazione, osservata di anno in anno in uno stesso luogo, è spesso dovuta ad una variazione dei parametri climatici; temperatura, vento, insolazione e pioggia sono i principali fattori meteorologici che influenzano il trasporto del polline dalla sorgente, rappresentata dalla pianta, al luogo di arrivo.

Le concentrazioni polliniche sono generalmente alte in giorni caldi e soleggiati, basse in giorni freddi e piovosi. Similmente, basse concentrazioni si hanno con la sorgente pollinica bagnata a causa di una pioggia recente o per la nebbia, mentre alte concentrazioni si rilevano quando la sorgente è asciutta e riscaldata.

1.1 Aerobiologia

L'aerobiologia è una disciplina scientifica che studia la sorgente, la dispersione, il trasporto e la deposizione delle particelle di origine biologica presenti nell’atmosfera e il loro effetto nell'ambiente.

Le applicazioni in campo sanitario del monitoraggio aerobiologico sono molteplici, la più nota è sicuramente quella legata alle allergie. Un ulteriore aiuto per il controllo clinico della sintomatologia, sia per il medico che per il paziente allergico, è dato dalla possibilità di costruire, partendo dai dati, i calendari di pollinazione: la distribuzione temporale degli aero-allergeni (pollini e spore fungine) in un dato luogo.

2. I Dati

Di seguito sono riportate le tipologie dei dati ritenute utili per affrontare gli studi; i dati a disposizione arrivano fino all'8 Aprile 2012 e sono stati tutti raccolti da ARPAV.

Concentrazioni polliniche

ARPAV esegue il monitoraggio quotidiano dei pollini [unità di misura: granuli/m³] a partire dal 2001; queste concentrazioni derivano principalmente da dieci stazioni venete [vedi Tabella 1] e da nove specie [vedi Tabella 2].

| Provincia | Stazione |
|-----------|----------------|
| Belluno | Belluno centro |
| | Agordo |
| | Feltre |
| Padova | Padova centro |
| | Teolo |
| Rovigo | Rovigo centro |
| Treviso | Treviso centro |
| Venezia | Mestre |
| Verona | Verona centro |
| | Legnago |
| Vicenza | Vicenza centro |

Tabella 1: Stazioni di monitoraggio

| Specie | Cod_SPC |
|--------------|---------|
| Betulaceae | 501 |
| Compositae | 502 |
| Corylaceae | 503 |
| Fagaceae | 504 |
| Graminaceae | 505 |
| Oleaceae | 506 |
| Urticaceae | 508 |
| Cupressaceae | 509 |
| Salicaceae | 518 |

Tabella 2: Specie rilevate

Temperature e Precipitazioni

ARPAV raccoglie quotidianamente le temperature orarie dell'aria a 2 metri da terra [unità di misura: °c] e le precipitazioni [unità di misura: mm/m²] dall'anno 2002, relativi ad oltre 500 stazioni distribuite nel territorio veneto.

Radiazione Solare Globale

ARPAV esegue il monitoraggio quotidiano della radiazione solare globale. Si tratta del valore ottenuto dalla somma della radiazione solare diretta e della radiazione globale diffusa, ricevuta dall'unità di superficie orizzontale. L'unità di misura utilizzata per i valori medi che si registrano a scala giornaliera è MJ/m².

2.1 Somma Termica

Per inserire la temperatura in un modello predittivo gli esperti ARPAV hanno consigliato una sua trasformazione lineare: la somma termica. Questo permette di tener conto sia di uno storico che subisce la pianta che della relazione che si è verificato esserci tra la stagione pollinica e il vegetale.

Durante il ciclo vitale, le piante, richiedono una certa quota di calore (energia) per le varie fasi della loro crescita. La durata del ciclo risulta generalmente più breve quanto maggiore è la quantità di energia ricevuta che è a sua volta correlata con la temperatura media giornaliera. Si possono prevedere eventuali anticipi o ritardi di maturazione in funzione di andamenti stagionali più caldi o più freddi rispetto alla media.

Per “somma termica”, definita anche “Grown Degree Days” (gradi utili giorno), si intende la sommatoria giornaliera dei gradi di temperatura che consente di valutare l’andamento delle temperature durante la stagione vegetativa della pianta. Questo parametro si calcola come sommatoria delle differenze tra le temperature medie giornaliere e lo “zero di vegetazione” ossia quel valore di temperatura variabile per ogni specie di pianta sotto il quale i processi di sviluppo non si attivano.

In questo caso la formula per il calcolo della Somma Termica è:

$$(\text{Temp. Media} - \text{zero vegetazione}) = \text{GRADI UTILI GIORNO [GDD]}$$

$$\text{S.T.} = \text{Somma(GDD)}$$

Se la temperatura media giornaliera è uguale o minore allo zero di vegetazione, non si accumulano gradi utili giorno.

In termini pratici se la somma termica è maggiore di zero significa che la temperatura è stata mediamente maggiore di quella richiesta per la crescita ottimale della specie, mentre se la somma termica è minore di zero allora la temperatura è stata troppo bassa.

3. La Strumentazione

Le concentrazioni polliniche sono state raccolte quotidianamente tramite il captaspore (catturatore pollinico). Questo strumento consiste in un aspiratore a portata continua e costante che indirizza il getto d'aria verso un tamburo rotante sul quale è stesa una pellicola adesiva così da poter catturare le particelle biologiche presenti nell'aria ottenendo le concentrazioni [unità di misura: granuli/m³ d'aria] di polline che si andranno a studiare.

Il tamburo rotante effettua un giro completo ogni settimana quindi cadenzialmente è necessario cambiare la pellicola adesiva. Quest'ultima viene preparata spalmando un sottile strato di silicone trasparente su di un foglio di plastica altrettanto trasparente (per facilitare la successiva valutazione al microscopio). Questo nastro avvolto sul tamburo risulta diviso in 7 aree che corrispondono all'arco di rotazione effettuato dal tamburo in una settimana.

Il flusso d'aria monitorato è costante e pari a 10 litri al minuto, simulando così il respiro di un essere umano. Contiene le eventuali spore che andranno ad aderire allo strato di silicone e trattenute fino al momento in cui non verrà estratto e portato in laboratorio.

All'interno della camera sterile la pellicola viene tagliata nei 7 giorni e trattata in modo da colorare i pollini presenti nella striscia per distinguerli dagli altri elementi atmosferici come le polveri.

Successivamente viene racchiusa in un vetrino e studiata al microscopio ottico per il conteggio quali-quantitativo dei pollini.

Il metodo di lettura è di tipo statistico: viene letta solo una opportuna parte della superficie, chiamata campione, che verrà poi rapportata all'intera parte del vetrino.

Obiettivi della rete di monitoraggio:

- Riconoscere il tipo di polline;
- Quantificarne la presenza.

4. Metodologia ed Elaborazione

La fase iniziale dello stage è servita per conoscere e comprendere l'oggetto in studio attraverso il materiale ricevuto dall'azienda, ricerche nel web e incontri con specialisti ARPAV del settore. Ciò ha portato a identificare alcuni parametri descrittivi utili all'azienda per studiare il fenomeno.

Parametri descrittivi della Stagione Pollinica (SP):

Indice pollinico annuo [Cum]:

si tratta della somma delle concentrazioni giornaliere annue, cumulata.

Data di inizio della SP [5%]:

giorno in cui la somma cumulata della concentrazione giornaliera raggiunge il 5% rispetto la concentrazione totale annua.

Data di picco [Picco]:

giorno in cui si registra il più elevato valore di concentrazione.

Data di massima fioritura [50%]:

giorno in cui la somma cumulata della concentrazione giornaliera raggiunge il 50% rispetto la concentrazione totale annua.

Data di fine SP [95%]:

giorno in cui la somma cumulata della concentrazione giornaliera raggiunge il 95% rispetto la concentrazione totale annua.

Durata della SP [Durata]:

numero di giorni compresi tra la data di inizio e di fine della SP.

[Validazione]:

percentuale dei dati presenti in una SP. Questo parametro viene utilizzato per pesare le rilevazioni [in teoria andrebbero rifiutate le concentrazioni annue i cui valori di validazione sono inferiori all'85%].

Le concentrazioni di polline rilevate si trovano in un DataBase Access [ARPAV/DB/DBPollini.mdb] all'interno della table [concentraz] avente i seguenti campi:

- [Cod_STD] identificativo della stazione di monitoraggio.
- [Cod_SPC] identificativo della specie analizzata.
- [Data_mis] la data a cui corrisponde la rilevazione.
- [Concentraz] il numero di granuli pollinici rilevati.

A partire da questa tabella i dati sono stati estratti, con opportune query, e suddivisi in file [ARPAV/DATI/Concentrazioni] a seconda della specie e catalogato ciascuno per data e stazione.

E' opportuno far notare che si è scelto di non tenere i valori di tutti i giorni dell'anno ma solo della parte significativa, evitando così di occupare spazio con valori nulli o mancanti.

Le serie ottenute sono discontinue, in corrispondenza di alcune date mancano i valori delle concentrazioni; è stata assegnata la stringa "NA" (Not Available). L'assenza della rilevazione può essere dovuta ad un valore raccolto e giudicato errato, ad un guasto dello strumento oppure per la chiusura momentanea della stazione pollinica.

Esempio di un file:

| Data_mis | BL01 | BL03 | FE01 | PD01 | RO01 | TV01 | VE01 | VI01 | VR01 | VR02 | Totale |
|------------|------|------|------|------|------|------|------|------|------|------|--------|
| 16/03/2004 | 571 | NA | 318 | 1 | 12 | 45 | 51 | 1 | 19 | NA | 1018 |

Una volta catalogati i pollini nei file di testo txt sono state create le funzioni, cercando di renderle più generiche possibili, per eseguire l'analisi descrittiva utilizzando il software R-cran [ARPAV/Elaborazioni].

- **[PD]** prende in ingresso il file di una specie e restituisce gli indici dei Parametri Descrittivi rispetto ad ogni stazione e anno.
- **[GR1]** prende in ingresso l'output di [PD] e restituisce i Grafici, divisi per anno, degli indici temporali delle stazioni [5%, 50%, 95%]. Permette il confronto della stagione pollinica, a parità di anno e specie, del comportamento nelle diverse stazioni venete.
- **[GR2]** prende in ingresso l'output di [PD] e restituisce i Grafici a barra divisi per anno del totale delle concentrazioni polliniche osservate nelle diverse stazioni. Permette il confronto quantitativo del polline rispetto le stazioni a parità di anno e di specie.
- **[M]** visualizza le medie nei diversi anni delle stazioni rispetto gli indici della funzione [PD] pesate con la validazione.

A seguito di questa grande quantità di informazioni si è scelto di formalizzare il tutto per un'unica specie di polline: La Graminacea.

5. Scheda tecnica della Graminacea

Graminacea - Fam. Graminaceae [cod_SPC 505]

«Piante erbacee con fioritura in primavera-estate e fiori riuniti in spighe che si sviluppano in ogni zona climatica. La famiglia comprende i cereali ad uso alimentare e numerose piante spontanee molto resistenti, per un totale di 5000 specie vegetali differenti.»



Immagine 1: Piante Graminacee.

Pianta: Si tratta della famiglia principale dell'ordine della Glumiflore e di essa sono fatte le più importanti culture strategiche mondiali, dalle quali si ricavano: farine, cereali, e alimenti elaborati quali pasta, pane e bevande alcoliche.

Habitat: Gli ambienti preferenziali di crescita sono quelli continentali con aridità frequente e temperature estreme come le steppe o praterie. Sono comunque presenti nella regione mediterranea, allo stato spontaneo occupando praticamente tutti i tipi di habitat, dai boschi ai luoghi umidi.

Forma vegetativa: piante erbacee con fusti cilindrici cavi articolati in internodi.

Foglie: si presentano strette e allungate; possiedono una parte inferiore avvolta attorno al fusto ed una parte superiore libera.

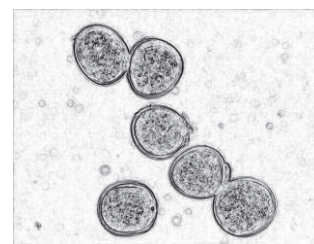
Fiori: le graminacee hanno fiori organizzati in piccole infiorescenze dette spighe, non sono colorati dato che non hanno bisogno di attrarre gli insetti per l'impollinazione. I fiori non hanno petali e quando il polline è pronto viene semplicemente portato via dal vento.

Fioritura: nella nostra regione inizia in primavera e si protrae per tutta l'estate.

Polline: il granulo pollinico ha forma sferoidale e presenta una sola apertura circolare, il poro. La parete si ispessisce attorno al poro formando un anello.

Allergenicità: molto alta.

Immagine 2: Granuli di polline del *Triticum dicoccoides* [Frumento], della famiglia Graminaceae.



6. Analisi del polline della Graminacea

Per iniziare l'analisi esplorativa dei dati relativi alla Graminacea, viene acquisito il relativo file 505.txt contenente le concentrazioni.

Dato che riportare i singoli esiti di tutte le stazioni risulta oneroso si è scelto di visualizzare solamente le tabelle e i grafici per le rilevazioni di Padova e delle medie di tutte le stazioni. Padova è stata scelta, non solo per l'interesse logistico, trovandosi nel territorio universitario, ma anche perché è una delle serie di dati più complete tra quelle a disposizione.

Codice R:

```
y05<-read.table("DATI/Concentrazioni/505.txt", header=TRUE)
plot(ts(y05[,5],start=c(2001,01),frequency=244),xlab="Anni",
ylab="Concentrazioni", main="Serie Temporale St. Padova Centro")
plot(ts(y05[,12]/10, start=c(2001,01),frequency=244), xlab="Anni",
ylab="Concentrazioni", main="Serie Temporale Media St.")
```

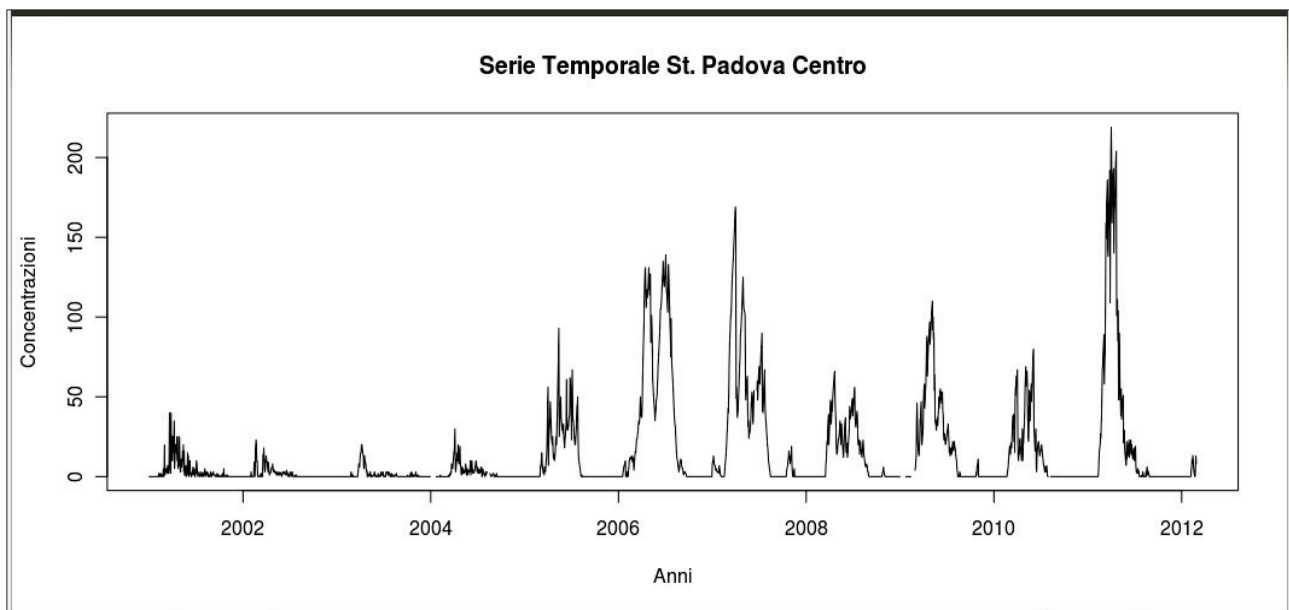


Grafico 1: Serie Temporale delle concentrazioni di polline delle Graminacee nella stazione di Padova Centro.

Si può notare che il grafico è abbastanza continuo, infatti si hanno pochi valori mancanti nella serie. I dati non sono relativi a tutti i giorni ma sono ridotti al periodo di interesse, avente in questo caso 244 giorni: dal primo marzo al 30 ottobre.

Osservando gli anni si nota una grande differenza tra le prime rilevazione e quelle successive. Ci si è posta la domanda di cosa avesse potuto causare questo forte aumento del polline monitorato.

Una delle ipotesi pensate è stata lo spostamento dello strumento di rilevazione dall'ospedale civile all'orto Botanico di Padova; ipotesi successivamente confermata.

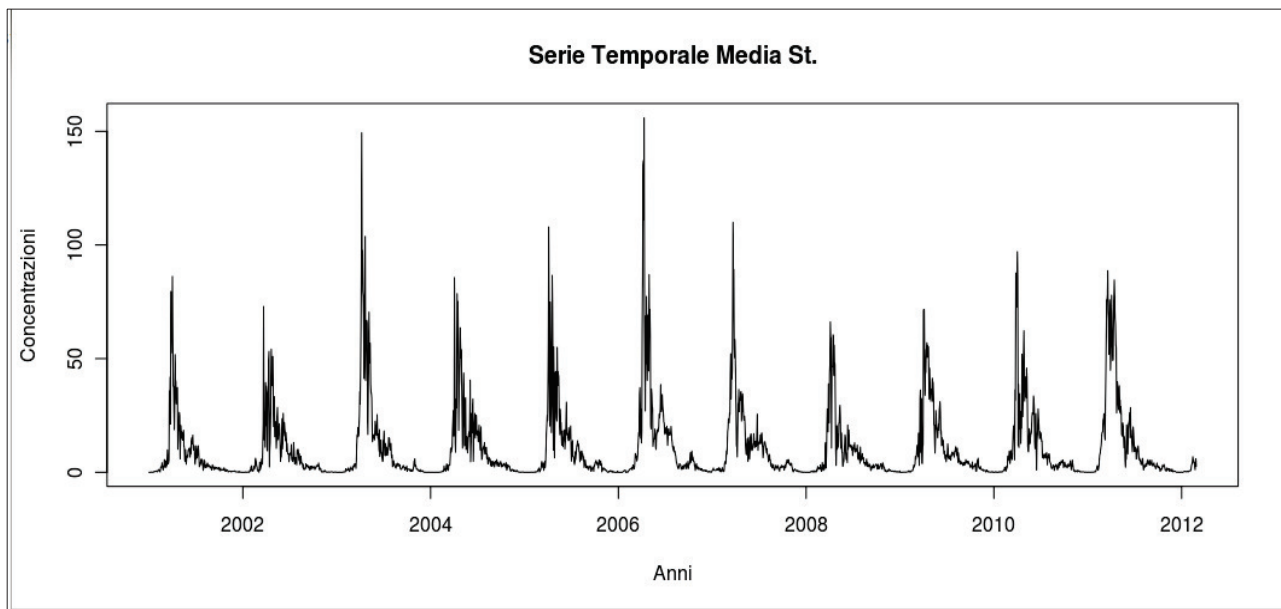


Grafico 2: Serie Temporale delle medie concentrazioni di polline delle Graminacee monitorato nelle stazioni.

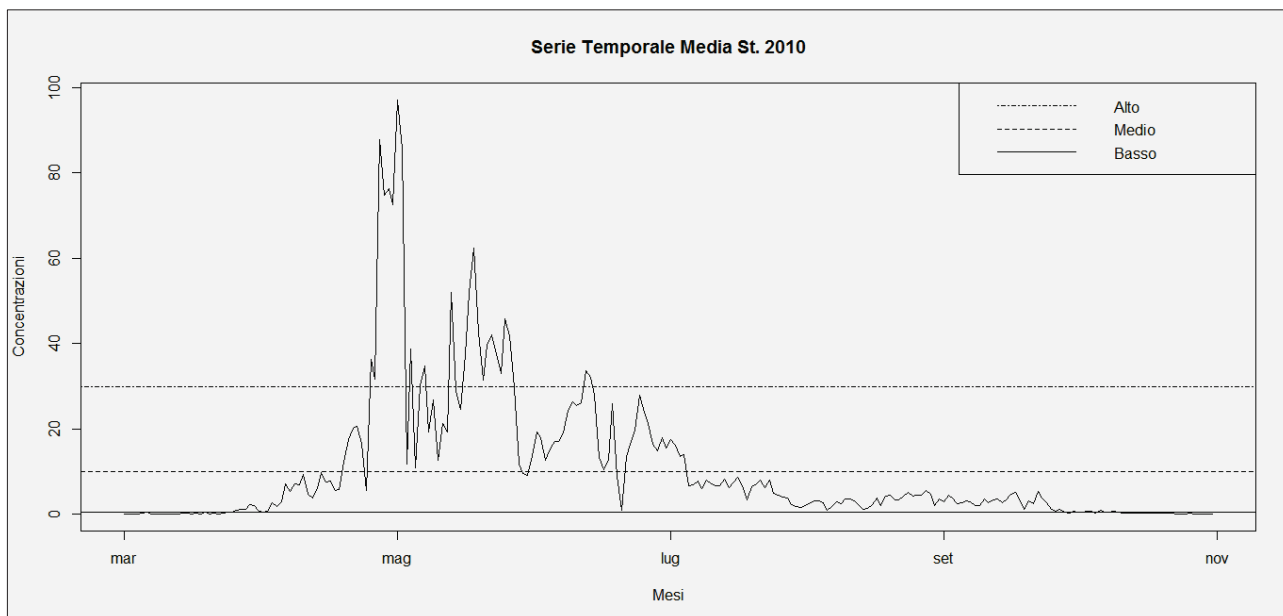


Grafico 3: Zoom del **Grafico 2** nell'anno 2010, con le soglie botaniche delle Graminacee.

Nel **Grafico 2** si nota che esiste una certa stagionalità negli anni: i vari punti critici, come il picco, rispettano le medie riportate all'ultima riga della **Tabella 3**. Infatti l'inizio della stagione pollinica è verso fine Marzo e poco dopo si verifica sempre il picco in corrispondenza ai primi giorni di Maggio. Negli'anni 2003 e 2006 son registrati i più alti livelli di concentrazioni osservate.

Il **Grafico 3** mostra in dettaglio la serie del 2010 e le fasi della stagionalità rispetto tutte le stazioni del Veneto. Contiene inoltre le quattro classi di concentrazione presenti nella tabella associate.

| Assente/molto basso | Basso | Medio | Alto |
|---------------------|------------|--------------|--------|
| 0 – 0.5 | > 0.5– 9.9 | > 9.9 – 29.9 | > 29.9 |

Tabella 3: Classi e intervalli di concentrazioni
[http://www.pollnet.it/valori_di_riferimento_it.asp]

Si sottolinea che le quattro classi di concentrazione **non corrispondono ai livelli di "rischio/allergia"**. Si tratta di una valutazione europea che fa riferimento alla quantità di polline presente nell'aria e non fornisce indicazioni sulle concentrazioni polliniche "soglia" scatenanti una reazione allergica.

Di seguito vengono riportate due tabelle: la prima contenente gli indici descrittivi di Padova, risultato della funzione [PD] e la seconda contenente i risultati della funzione [M], gli indici delle medie di tutte le stazioni.

| Inizio | Fine | Cum | 5% | 50% | 95% | Picco | Durata | Validità |
|------------|------------|--------|------------|------------|------------|------------|--------|----------|
| 26/03/2001 | 21/09/2001 | 862.0 | 03/04/2001 | 23/06/2001 | 12/09/2001 | 23/04/2001 | 180 | 100.00 |
| 22/03/2002 | 18/07/2002 | 402.0 | 28/03/2002 | 20/05/2002 | 13/07/2002 | 04/04/2002 | 119 | 100.00 |
| 07/04/2003 | 29/09/2003 | 347.0 | 24/04/2003 | 06/07/2003 | 21/09/2003 | 04/05/2003 | 176 | 100.00 |
| 26/03/2004 | 20/08/2004 | 561.0 | 16/04/2004 | 07/06/2004 | 13/08/2004 | 03/05/2004 | 148 | 91.39 |
| 11/04/2005 | 28/07/2005 | 2931.0 | 16/04/2005 | 04/06/2005 | 22/07/2005 | 29/05/2005 | 109 | 100.00 |
| 13/03/2006 | 24/08/2006 | 8906.0 | 23/03/2006 | 03/06/2006 | 15/08/2006 | 02/07/2006 | 165 | 100.00 |
| 01/03/2007 | 30/09/2007 | 7371.0 | 11/03/2007 | 15/06/2007 | 19/09/2007 | 30/04/2007 | 214 | 97.13 |
| 21/04/2008 | 21/09/2008 | 3119.0 | 28/04/2008 | 06/07/2008 | 14/09/2008 | 14/05/2008 | 154 | 100.00 |
| 09/04/2009 | 20/09/2009 | 4577.0 | 17/04/2009 | 30/06/2009 | 14/09/2009 | 24/05/2009 | 165 | 90.16 |
| 05/04/2010 | 18/07/2010 | 2893.0 | 10/04/2010 | 27/05/2010 | 12/07/2010 | 12/06/2010 | 105 | 97.95 |
| 29/03/2011 | 07/08/2011 | 7606.1 | 04/04/2011 | 02/06/2011 | 31/07/2011 | 01/05/2011 | 132 | 100.00 |
| 26/03/2012 | 08/04/2012 | 91.2 | 26/03/2012 | 01/04/2012 | 07/04/2012 | 30/03/2012 | 14 | 100.00 |

Tabella 4: Statistiche descrittive stazione di Padova

Nella **Tabella 4** risulta interessante notare come il picco si trovi molto vicino all'inizio della stagione pollinica, segno di una distribuzione asimmetrica.

| Stazione | MInizio | MFine | MConc | M5% | M50% | M95% | MPicco | MDurata |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|--------------|
| Belluno Centro | 29-mar | 16-set | 1867 | 06-apr | 26-giu | 09-set | 14-mag | 172 |
| Belluno Agordo | 05-apr | 29-ago | 824 | 12-apr | 17-giu | 22-ago | 06-mag | 147 |
| Belluno Feltre | 30-mar | 17-set | 2156 | 07-apr | 24-giu | 09-set | 10-mag | 172 |
| Padova Centro | 28-mar | 26-ago | 3604 | 07-apr | 11-giu | 18-ago | 12-mag | 151 |
| Rovigo Centro | 23-mar | 11-ott | 2238 | 03-apr | 02-lug | 02-ott | 03-mag | 203 |
| Treviso Centro | 18-mar | 15-ott | 2064 | 31-mar | 01-lug | 08-ott | 01-mag | 212 |
| Venezia Mestre | 16-mar | 20-ott | 2096 | 27-mar | 07-lug | 12-ott | 03-mag | 219 |
| Vicenza Centro | 17-mar | 13-ott | 5059 | 27-mar | 30-giu | 05-ott | 04-mag | 212 |
| Verona Centro | 13-mar | 26-ott | 4663 | 25-mar | 08-lug | 17-ott | 28-apr | 228 |
| Verona Legnago | 15-mar | 25-ott | 3076 | 27-mar | 06-lug | 15-ott | 01-mag | 225 |
| Veneto | 22-mar | 01-ott | 2764,7 | 01-apr | 28-giu | 23-set | 05-mag | 194,1 |

Tabella 5: Medie annue delle Statistiche Descrittive di tutte le stazioni.

La **Tabella 5** è stata creata ponderando gli indici annui delle stazioni con la validità del dato analizzato. L'ultima riga, Veneto, contiene la media delle stazioni, ciò permette di effettuare una valutazione su scala regionale per determinare, anno dopo anno, la stagione pollinica.

Nel Veneto la stagione pollinica delle Graminacee parte dal primo Aprile e termina la seconda metà di settembre con un picco verso i primi giorni di maggio. Queste informazioni collegate con un calendario di pollinazione risultano estremamente utili per le persone allergiche a questi pollini.

Una lettura sicuramente più diretta e un confronto facilitato di questa tabella viene dato dal seguente grafico.

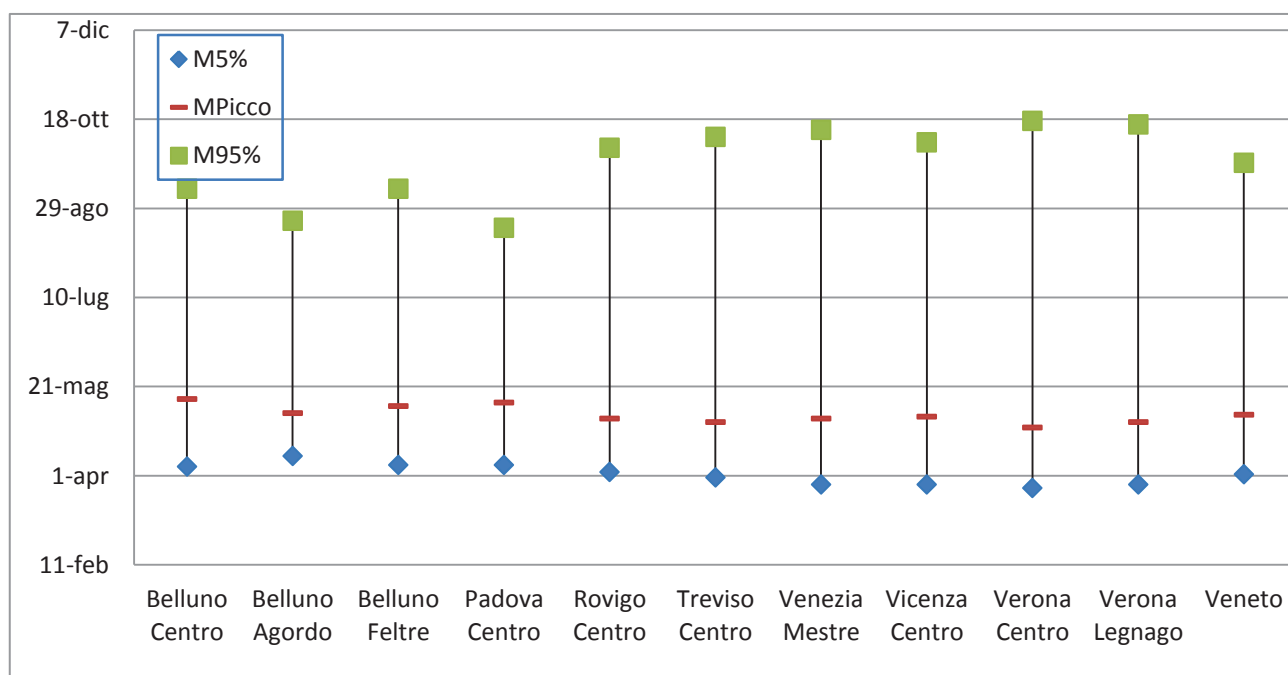


Grafico 4: Rappresentazione della tabella 5

Osservando il **grafico 4** si nota che le stagioni polliniche delle prime quattro stazioni hanno un andamento diverso rispetto le altre, infatti l'inizio ed il picco massimo è ritardato e finiscono oltre un mese prima.

Questo fenomeno è spunto per futuri argomenti di studio, per verificare se esiste un legame tra la concentrazione pollinica e la zona di rilevazione.

7. Studio sulla Somma Termica rispetto le Zone

La parte di modellazione dei dati è iniziata col cercare le possibili variabili che maggiormente influenzano la concentrazione pollinica; a tal fine si è svolto un incontro il 28 Maggio coinvolgendo più professionisti ARPAV. Mostrando le analisi ottenute è emersa una probabile relazione tra le somme termiche e il ciclo fenologico della pianta, ne segue perciò la possibile suddivisione del Veneto in 2 zone.



Immagine 3: suddivisione del Veneto

La Zona1 coinvolge la parte alta del Veneto, comprendente la provincia di Belluno e la zona dolomitica.

La Zona2 invece coinvolge la parte bassa del Veneto, con il territorio della pianura Padana.

Questa scelta di dividere il territorio è stata presa in base alle diverse caratteristiche climatiche e morfologiche delle due zone e dai risultati ottenuti dal **grafico 4**.

A questo punto è stato scelto di procedere con l'analisi della somma termica poiché da altri studi si sa che esiste una forte influenza con la fase vegetativa della pianta e si è analizzato questo evento sulle due zone per verificare la significatività dei due gruppi.

Lo scopo di questa seconda parte è di arrivare alla costruzione di un modello previsionale per determinare le fasi della stagione pollinica di una determinata specie conoscendo la zona di riferimento e la somma termica finora cumulata.

Essendo interessati ad uno studio sulla significatività dei due gruppi, sarebbe risultato superfluo sottrarre alla somma termica lo zero vegetativo, come invece viene richiesto dalla definizione. Inoltre le ricerche effettuate per trovarlo a livello specifico di ogni famiglia sono state vane. Per questi due motivi si è utilizzata una trasformazione della somma termica, ma per semplicità in seguito verrà ugualmente chiamata nello stesso modo.

7.1 Preparazione del Data-Set

Il procedimento per arrivare a tale scopo è stato quello di trovare le stazioni meteorologiche presenti nei dintorni delle stazioni polliniche, possibilmente entro un raggio di 10 km (distanza entro la quale si spargono le particelle polliniche).

Utilizzando l'applicazione web ArcGis [<http://www.arcgis.com/>] sono state inserite le coordinate di tutte le stazioni nella cartina geografica del Veneto, per individuare agevolmente le stazioni meteorologiche di nostro interesse, circa 4 per ogni stazione pollinica in modo da coprire i punti cardinali.

Per ognuna di queste stazioni sono stati ricavati i dati, tramite query poste al database meteo di ARPAV, ottenendo così dei file denominati con [codice stazione meteo]_[codice stazione pollinica] aventi la relativa tabella:

- Data [giorno a cui appartengono le osservazioni];
- Precipitazione [quantità giornaliera in mm di pioggia];
- Temperatura [media giornaliera in gradi centigradi].

Esempio di un file: 167_PD01

| Data | Precipitazione | Temperatura |
|------------|----------------|-------------|
| 07/06/2007 | 0.13 | 20.04 |
| 08/06/2007 | 0.03 | 21.56 |

Questi file sono stati posizionati in due cartelle differenti a seconda che appartenessero alla zona 1 o 2 e, sempre utilizzando il programma R-

cran, sono state create le funzioni:

- **[ST]** prende in ingresso l'output di [PD] di una specie e restituisce le somme termiche divise per zona. E' la sommatoria delle temperature che partono dal 1 Gennaio fino alla media annua di inizio della stagione pollinica (indice descrittivo: 5%) della zona di riferimento.

Le due tabelle in uscita, **zona1** e **zona2**, avranno nelle colonne le stazioni meteorologiche che sono state scelte, nelle righe gli anni dal 2001 al 2012 e all'interno le somme termiche.

- **[Dati]** prende in ingresso le due tabelle create da [ST] e restituisce il data-set, privato dagli "NA", che servirà per fare le analisi, avente nelle righe le osservazioni e nelle colonne:
 - **Data_Media_5%:** contiene la data media annua degli inizi delle stagioni polliniche [5%] calcolata sulle stazioni polliniche presenti nella stessa zona.
 - **Somme_Termiche:** sommatoria della temperatura di una determinata stazione climatica e di un determinato anno, parte dal 1 Gennaio fino alla Data_Media_5% dell'anno e della zona associati.
 - **Zona:** variabile binomiale a cui viene associato il valore 0 se l'osservazione appartiene alla zona1 (alta, dolomiti) altrimenti 1 se appartiene alla zona2 (bassa, pianura).
 - **Stazione_Clima:** è formata da due codici univoci, determinano le stazioni climatica e pollinica di riferimento [esempio: 167_PD01 è la stazione climatica 167 e la stazione pollinica di Padova].
 - **Anno:** riporta l'anno a cui appartiene l'osservazione. Si noti che tutte le osservazioni del 2001 sono state eliminate, a differenza delle concentrazioni polliniche i dati climatici partono dal 2002.

| Data_Media_5% | Somme_Termiche | Zona | Stazione_Clima | Anno |
|---------------|----------------|------|----------------|------|
| 07/04/2002 | 223.7 | 0 | 13_BL03 | 2002 |
| 01/04/2003 | 75.27 | 0 | 13_BL03 | 2003 |
| 27/04/2004 | 240.98 | 0 | 13_BL03 | 2004 |
| 18/04/2005 | 87.84 | 0 | 13_BL03 | 2005 |
| 18/04/2006 | 62.22 | 0 | 13_BL03 | 2006 |
| 21/03/2007 | 208.08 | 0 | 13_BL03 | 2007 |
| 08/04/2008 | 187.38 | 0 | 13_BL03 | 2008 |
| 12/04/2009 | 112.9 | 0 | 13_BL03 | 2009 |
| 19/04/2010 | 110.36 | 0 | 13_BL03 | 2010 |
| 07/04/2002 | 153.31 | 0 | 19_BL03 | 2002 |
| 01/04/2003 | -5.91 | 0 | 19_BL03 | 2003 |

Tabella 6: Prime 11 righe del data-set delle somme termiche, il codice viene riportato di seguito.

Codice R:

```
y05<-read.table("DATI/Concentrazioni/505.txt", header=TRUE)
S<-ST(PD(y05), 0)
dati<-DATI(S)
dati[11,]
```

7.2 Analisi

Per iniziare le analisi esploriamo il Data-Set che è stato restituito dalla funzione [DATI] guardando la numerosità, alcune statistiche descrittive di sintesi dei due diversi gruppi, la correlazione e la distribuzione dei dati.

Codice R:

```
n<-table(dati$zona)
```

| Zona | Dummy | Numerosità |
|------|-------|------------|
| 1 | 0 | 104 |
| 2 | 1 | 231 |

```
summary(dati[dati$zona==0,1])
```

```
summary(dati[dati$zona==1,1])
```

| zona | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|----------|---------|---------|---------|---------|---------|
| 1 | -148.500 | 5.558 | 151.000 | 153.300 | 245.100 | 561.200 |
| 2 | 0.0 | 397.5 | 446.9 | 454.4 | 517.6 | 727.3 |

```
cor(dati[,1:2])
```

| | somme_t | zona |
|---------|---------|------|
| somme_t | 1 | 0.73 |
| zona | 0.73 | 1 |

```
boxplot(dati[dati$zona==0,1],dati[dati$zona==1,1], main="Boxplot
Zone", ylab="SommeTermiche", xlab="Zone",xaxt="n")
```

```
axis(1,at=1:2, lab=c("Zona1","Zona2"))
```

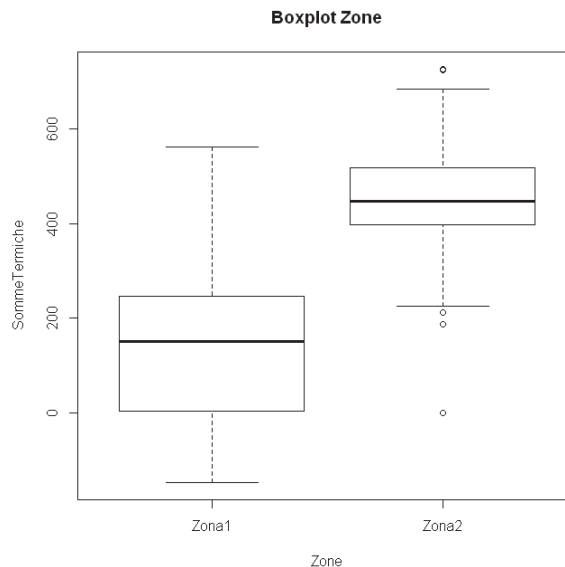


Grafico 5: boxplot delle distribuzioni delle due zone.

Avendo un numero superiore di stazioni polliniche appartenenti alla zona bassa si hanno molte più osservazioni (più del doppio) nel secondo gruppo. Dal **grafico 5** si notano le distribuzioni dei due gruppi:

- non risultano normali come ci si sarebbe aspettato, infatti le osservazioni non sono indipendenti tra loro;
- sono diverse in quanto sembrano avere una variabilità e soprattutto una media differenti, queste ipotesi saranno confermate dai test che successivamente faremo.

La Zona1 ha una variabilità eccessiva e una distribuzione fortemente asimmetrica a destra. Per questo motivo sono state fatte delle indagini per controllati meglio questi dati.

```
sum(dati[dati$zona==0,1]>400)
```

```
[1] 9
```

Queste 9 osservazioni corrispondono tutte alla stazione 245_FE01, situata a Quero, località molto vicino al confine delle due zone ed effettivamente ha valori di temperatura elevati rispetto al suo gruppo di appartenenza. Per questo motivo si è scelto di escluderla dall'analisi in quanto dato anomalo.

Andando a ripetere gli studi si ottengono buoni cambiamenti nella media:

```
mean(dati[dati$zona==0,1])
```

```
[1] 153.2695 # media con la stazione di Quero
```

```
mean(dati[dati$zona==0,1][~c(54:63)])
```

```
[1] 123.4803      # media senza la stazione di Quero
```

```
dati<-dati[-c(54:63),]
```

```
table(dati$zona)
```

| | Dummy | Numerosità |
|-------|-------|------------|
| zona1 | 0 | 94 |
| zona2 | 1 | 231 |

```
summary(dati[dati$zona==0,1])
```

```
summary(dati[dati$zona==1,1])
```

| zona | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|------|----------|---------|---------|---------|---------|---------|
| 1 | -148.500 | 0.00 | 111.600 | 123.500 | 211.900 | 477.00 |
| 2 | 0.0 | 397.500 | 446.900 | 454.400 | 517.600 | 727.300 |

```
cor(dati[,1:2])
```

| | somme_t | zona |
|---------|---------|------|
| somme_t | 1 | 0.78 |
| zona | 0.78 | 1 |

```
boxplot(dati[dati$zona==0,1],dati[dati$zona==1,1], main="Boxplot
Zone", ylab="SommeTermiche", xlab="Zone",xaxt="n")
axis(1,at=1:2, lab=c("Zona1","Zona2"))
```

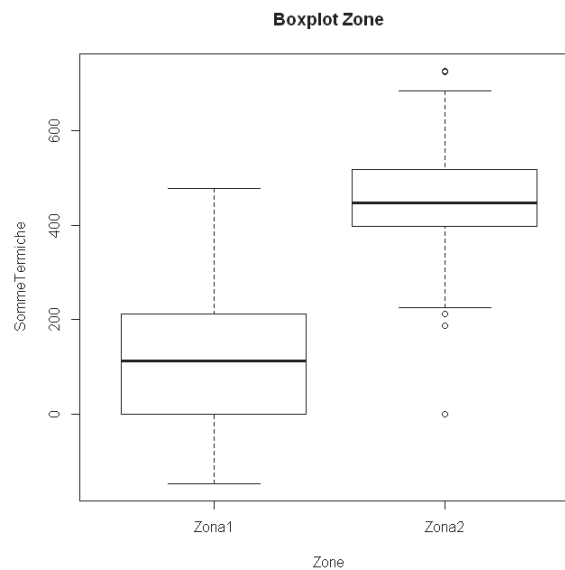


Grafico 6: boxplot delle distribuzione delle due zone senza la stazione di Quero.

Confrontando i due grafici a baffi si osserva una diminuzione della varianza per la Zona 1. Questo miglioramento si nota anche nella matrice di correlazione sopra riportata. Verifichiamo con dei semplici test se le ipotesi pensate precedentemente guardando i grafici sono corrette:

```
var.test(dati[dati$zona==0,1],dati[dati$zona==1,1])
```

F test to compare two variances

data: dati[dati\$zona == 0, 1] and dati[dati\$zona == 1, 1]

F = 1.7981, num df = 93, denom df = 230, p-value = 0.0004351

alternative hypothesis: true ratio of variances is not equal to 1

95% confidence interval: 1.293043 2.563943

sample estimates:ratio of variances 1.798061

Il var.test permette di confrontare le varianze di due popolazioni con l'obiettivo di capire se sono uguali. Essendo il livello di significatività osservato molto basso [0.0004351], viene rifiutata l'ipotesi nulla, H_0 : il rapporto delle due varianze è 1, e si accetta l'ipotesi alternativa, H_1 : il rapporto delle varianze dei due gruppi è diversa da 1. Dunque i due gruppi hanno una differente varianza: sono eteroschedastici.

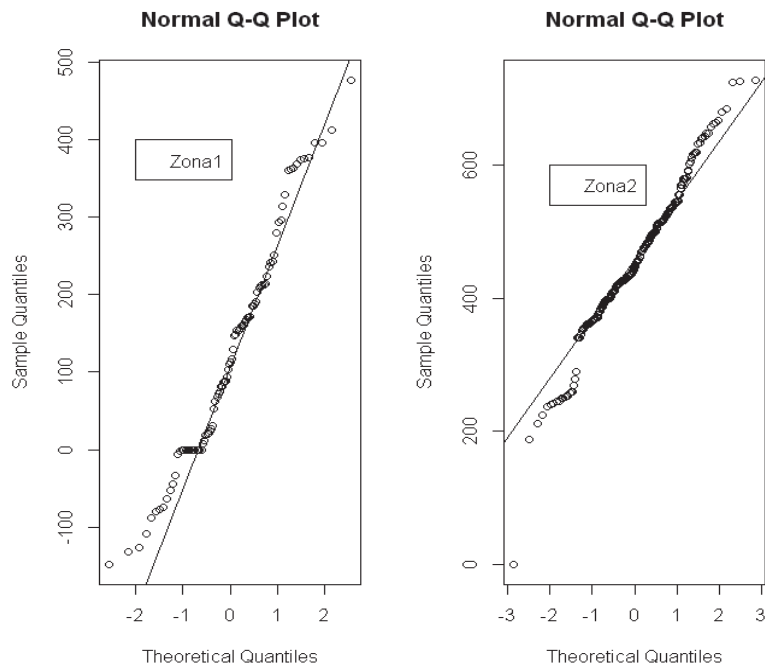


Grafico 7: verifica normalità delle due zone.

```
shapiro.test(dati[dati$zona ==0,1])
```

Shapiro-Wilk normality test

W = 0.9689, p-value = 0.02451

```
shapiro.test(dati[dati$zona ==1,1])
```

Shapiro-Wilk normality test

W = 0.9787, p-value = 0.001498

I due test di shapiro-Wilk testano la normalità dei due gruppi, i due livelli di significatività portano a rifiutare l'ipotesi nulla, H0: i dati seguono una distribuzione normale. Anche il **grafico 7** mostrano che le code delle due distribuzioni non sono propriamente normali.

```
t.test(dati[dati$zona ==0,1], dati[dati$zona ==1,1])
```

Welch Two Sample t-test

t = -19.9953, df = 137.02, p-value < 2.2e-16

alternative hypothesis: true difference in means is not equal to 0

95% confidence interval: -363.6929 -298.2321

sample estimates: mean of x mean of y 123.4803 454.4428

Il t.test invece saggia l'ipotesi che la differenza delle medie dei due campioni sia nulla contro l'ipotesi alternativa che sia diversa da zero. Il livello di significatività riportato è molto basso e porta al rifiuto dell'ipotesi nulla.

Per concludere i due gruppi derivano da distribuzioni significativamente diverse per media e varianza.

7.3 Modellazione

Questo paragrafo propone uno studio sui modelli di classificazione delle osservazioni nelle due zone conoscendo le somme termiche. Si effettueranno confronti tra un modello semplice [m0], avente solamente le somme termiche e la zona, e quello completo [m1] avente 11 variabili dummy utilizzate per identificare l'anno.

La variabile dell'anno 2001 è una costante di zeri perché non si dispone delle osservazioni. L'anno 2012 è la variabile dummy esclusa per evitare un problema di collinearità: presenza di una forte correlazione tra le variabili esplicative, quindi non fornisce delle informazioni aggiuntive sulla variabile risposta. Si ottiene così questa nuova tabella dei dati:

| Data_Media_5% | Somme_T | Zona | Stazione_Clima | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 | 2011 |
|---------------|---------|------|----------------|------|------|------|------|------|------|------|------|------|------|
| 07/04/2002 | 223.7 | 0 | 13_BL03 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 01/04/2003 | 75.27 | 0 | 13_BL03 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 27/04/2004 | 240.98 | 0 | 13_BL03 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18/04/2005 | 87.84 | 0 | 13_BL03 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 18/04/2006 | 62.22 | 0 | 13_BL03 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 21/03/2007 | 208.08 | 0 | 13_BL03 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Ad ogni modello viene poi ottenuta la tabella di contingenza [tab] e su di essa vengono calcolati gli errori di classificazione:

Errore Assoluto => $1 - [\text{sum}(\text{diag}(\text{tab})) / \text{sum}(\text{tab})]$;

Errore Relativo => $\text{Errore Assoluto} / (1 - [\text{max}(\text{marg}) / \text{numerosità totale}])$.

Quest'ultimo errore è utile per confrontare i modelli di diversa entità.

QDA - ANALISI DISCRIMINANTE QUADRATICA

Per utilizzare questo metodo si necessita di gruppi normali, pur non avendoli nel nostro caso gli assumiamo asintoticamente tali.

```
m0<-valinc1.qda(zona~somme_t, campione=dati)
errori(m0)
```

| | | |
|---|----|-----|
| | 0 | 1 |
| 0 | 78 | 16 |
| 1 | 17 | 214 |

Errore Assoluto= 0.128

Errore Relativo= 0.414


```
m1<-valinc1.qda(zona~., campione=dati)
errori(m1)
```

| | | |
|---|----|-----|
| | 0 | 1 |
| 0 | 82 | 12 |
| 1 | 4 | 227 |

Errore Assoluto= 0.066

Errore Relativo= 0.212

In questo modello, come ci si aspettava aumentando le variabili, la stagionalità degli anni migliora la capacità predittiva.

GLM – MODELLO LINEARE GENERALIZZATO

Avendo la variabile risposta binaria utilizziamo la famiglia binomiale per creare due modelli lineari generalizzati, uno semplice e l'altro complesso e testiamo il migliore con il test anova.

```
m1_glm<-glm(zona~.,data=dati,family = binomial(link = logit))
m0_glm<-glm(zona~somme_t,data=dati,family = binomial(link = logit))
anova(fit0_glm,fit_glm, test="Chisq")
```

Analysis of Deviance Table

Model 1: zona ~ somme_t

Model 2: zona ~ somme_t + anno2002 + anno2003 + anno2004 + anno2005 + anno2006 + anno2007 + anno2008 + anno2009 + anno2010 + anno2011

| | Resid. Df | Resid. Dev | Df | Deviance | Pr(>Chi) |
|---|-----------|------------|----|----------|-----------|
| 1 | 333 | 199.78 | | | |
| 2 | 323 | 143.80 | 10 | 55.974 | 2.076e-08 |

Il test ci porta al rifiuto dell'uguaglianza del modello semplice con quello complesso. Conviene utilizzare un modello che tenga conto degli anni perché porta una maggiore informazione.

Le analisi effettuate si fermano qua, un punto dove un'altra persona potrà in seguito continuarle fino ad ottenere il modello cercato. Da questi studi sono emerse le due zone significative di divisione del territorio veneto per quanto riguarda la stagione pollinica e il tentativo di spiegarla attraverso le somme termiche una trasformazione della temperatura.

Oltre al lavoro iniziale per ottenere quelle funzioni in grado di descrivere una serie temporale sulle concentrazioni polliniche.

9. Altre attività svolte durante lo stage

9.1 Open Data

Con il termine Open Data ci si riferisce a quei dati che possono essere liberamente utilizzati e ridistribuiti da chiunque; sono soggetti alla necessità di citarne la fonte e, se condivisi, va mantenuta lo stesso tipo di licenza con cui sono stati originariamente rilasciati.

Requisiti che i dati devono possedere per essere definiti “aperti”:

A livello internazionale il riferimento più importante è il Manuale degli Open Data creato dalla Open Knowledge Foundation, dal quale si riassumono tre requisiti:

I. Disponibilità e Accesso:

i dati devono avere un formato modificabile e per un prezzo non superiore al costo di riproduzione, preferibilmente mediante download da Internet.

II. Riutilizzo e Ridistribuzione:

i dati devono consentire il riutilizzo e la ridistribuzione, ciò comprende la possibilità di combinarli con altri DB.

III. Partecipazione universale:

tutti devono essere in grado di usare, riutilizzare e ridistribuire i dati. Non devono essere poste discriminazioni in ambiti di iniziativa in riferimento a soggetti o gruppi. Per esempio, il divieto di utilizzare i dati per scopi commerciali o le restrizioni che permettono l'uso solo per determinati fini (quale quello educativo) non sono contemplabili.

Il permesso di utilizzare i dati anche a fini commerciali, rappresenta l'aspetto più importante di questo modello per il suo potenziale effetto positivo sull'avvio di nuove attività economiche. Le aziende private e i liberi professionisti esperti del settore, infatti, possono trarre dei vantaggi economici realizzando, a partire da dati liberi, dei prodotti e dei servizi a pagamento per i cittadini. Nell'ottobre 2010 è nata la licenza italiana per gli Open Data (IODL – Italian Open Data Licence), con lo scopo di permettere alle pubbliche amministrazioni italiane di diffondere i dati.

Open Data come strategia di sviluppo verso l'Open Government

Per “Open Government” si intende il concetto di Governance basato su strumenti e tecnologie che consentono alle amministrazioni di essere “trasparenti” nei confronti dei cittadini. In particolare prevede che tutte le attività dei governi e delle amministrazioni dello stato debbano essere disponibili, al fine di favorire azioni efficaci e garantire un controllo pubblico sull'operato.

Il formato “libero” dei dati pubblici rappresenta il presupposto di base affinché possa svilupparsi una vera e propria rete collaborativa e partecipativa tra le istituzioni e la comunità dei cittadini, sulle scelte di governo. Mediante queste strategie i cittadini non sono più consumatori passivi di informazioni messe a disposizione dai governi, ma hanno l’opportunità di arrivare a sviluppare veri e propri servizi e applicazioni a disposizione dell’intera comunità che vanno ad affiancarsi a quelli creati dalle istituzioni.

Per convincere un’organizzazione che il dato pubblico dovrebbe essere, nei limiti della tutela della privacy, “geneticamente aperto” e accessibile, si basti pensare che è stato finanziato dalla collettività e ad essa deve essere naturalmente restituito.

Veneto Big Data

Venerdì 15 Giugno 2012 si è tenuto un convegno a Mestre, al parco Scientifico Tecnologico Vega, denominato: Veneto Big Data. La giornata è stata suddivisa in due fasi:

1. durante la mattina è stato spiegato, tramite una “tavola rotonda”, cosa significa fare open-data, concetti che verranno riportati in seguito.

Lo spunto di maggiore interesse è stato il concetto di interoperabilità: gli enti raccolgono i dati e la comunità esprime la sua creatività creando applicazioni con essi, unendo così il pubblico con il privato per ottenere un migliore sviluppo.

2. nel pomeriggio c’era la possibilità di partecipare a tavoli ristretti che trattavano concetti più precisi e tangibili. La mia scelta è ricaduta sul tavolo “Open Data e Smart Cities” coordinato da Michele Vianello, direttore del complesso che ci ospitava.

La discussione trattava come migliorare il concetto di città per renderla più intelligente: la città che narra se stessa. Si è discusso su dove e come conservare i dati, espanderli attraverso piattaforme “social” rendendo possibile la comunicazione diretta tra cittadino e governo.

9.2. Noise-Tube

Si tratta di un progetto sviluppato in numerose città basato su tecnologia open-source; è una piattaforma collaborativa per il monitoraggio dell'inquinamento acustico utilizzando tecnologia mobile. In pratica i dispositivi portatili sono dei sensori che raccolgono il livello del rumore misurato, in decibel, dai singoli cittadini, i quali poi rendono disponibile l'informazione georiferita su una mappa collaborativa attraverso il web. Di seguito vengono riportati alcuni termini tecnici:

City Sensing tecnologia urbana che trasmette in rete, mappare i fenomeni che accadono in un territorio (DB geografico).

Sensor Web cittadini che raccontano il loro territorio attraverso mezzi meccanici e percettivi.

Mems (Micro Electro Mechanical Sistem) piccoli aggeggi uniti a dispositivi portatili per sfruttare microfoni.

Citizen Science i cittadini si fanno scienziati.

Vantaggi:

- misure diffuse e numerose;
- sensibilizzare i cittadini all'inquinamento acustico e al monitoraggio;
- basso costo unitario;
- i decibel vissuti quotidianamente da una persona sono difficilmente assimilabili da uno strumento statico;
- dati ottenuti quasi in tempo reale.

Limiti

- sensori a basso costo implicano maggiori errori strumentale ;
- persone non qualificate implicano maggiori errori umani;
- troppi dati, richiedono un metodo statistico per analizzarli;
- dati presi con dispositivi differenti risultano eterogenei tra loro, devono essere calibrati.

9.3 Inquinamento Acustico

Si è tenuta, il 9 Maggio 2012, una riunione coordinata da Rina Camporese, impiegata Istat, la quale ha scritto una tesi sul Noise Tube. Oltre a noi stagisti erano presenti alcuni esperti ARPAV nel monitoraggio del suono e il dirigente Luca Menini.

La riunione vergeva su questa nuova tecnica che sta prendendo il via in diversi paesi; dopo una breve panoramica sulla piattaforma si è discusso sulla qualità delle rilevazioni riportando delle osservazioni raccolte a Padova con queste tecniche confrontate con quelle monitorate da ARPAV. Di seguito vengono riportati i problemi riscontrati durante il confronto e le proposte per evitarli.

Errori di misurazione: alcuni studi affrontati dall'ARPAV hanno riportato che nelle singole osservazioni questi mezzi non sono paragonabili con i loro, tuttavia in un arco temporale gli errori si compensano e le osservazioni dopo esser state sottoposte a controlli possono essere divulgate.

Persone inesperte: è possibile ovviare alla scarsa preparazione del personale mediante dei tutorial. Alcuni esempi nel web hanno dato buoni risultati [esempi: open street e galaxy-zoo]. L'importante è spiegare tutto, come il controllo dello stato dello strumento prima di utilizzarlo, non mettere la mano sul microfono ecc. non dando nulla per scontato

Il problema resta il come coinvolgere la gente a fare queste osservazioni. Questo probabilmente è l'unico vero limite a cui si deve trovare un rimedio che invogli i cittadini.

11. Corso di PHP

All'interno dell'esperienza stage c'è stato proposto di seguire un corso di tre giorni con esame finale sulle basi di programmazione in PHP nato per i dipendenti ARPAV. Ritenendolo utile per la formazione di un informatico vi ho partecipato sebbene il primo giorno è stato un semplice ripasso dei concetti basilari della programmazione.

Interessante invece è stata la parte pratica di creazione di un menu per un eventuale sito in PHP e assistere al clima studentesco all'interno di un ente lavorativo.

12. Conclusioni e Ringraziamenti

Il risultato di questa esperienza è stato positivo sia per me sia per le persone che mi hanno seguito, all'interno dell'ufficio si era creato un clima di collaborazione. Ho conosciuto nuove realtà di lavoro dove uno statistico può applicarsi al meglio.

Per quanto riguarda gli obiettivi, erano ambiziosi e come si prevedeva non sono riuscito a concludere il modello, tuttavia sono stato d'aiuto nel trovare una prima variabile esplicativa: la somma termica. Sono rimasto soddisfatto dei risultati ottenuti e avrei avuto piacere poter continuare questi studi puntando poi sul creare un'applicazione web e pubblicando i risultati con una licenza libera.

Per concludere ringrazio le persone che si sono adoperate per aiutarmi in questa esperienza di stage ma anche universitaria.

13. Sitologia e Bibliografia

- Diapositive e materiali ricevuti dall'ARPAV, sono presenti nella cartella [ARPAV/Materiale].
- ” Botanica oggi – ambiente, vegetazione, sviluppo ” di M.N. Forgiarini, S. Raggi, C. Casali; Calderini edagricole. Cap. 20 il Fiore, Cap. 22 Il Glumiflore
- <http://www.noisetube.net/>
- <http://www.galaxyzoo.org/>
- <http://www.open-street.eu/index.asp?l=it>
- <http://www.arpa.veneto.it/temi-ambientali/pollini7>
sito dell'azienda ospitante
- <http://www.polnet.com/>
- http://it.wikipedia.org/wiki/Dati_aperti
- <http://www.irreer.it/boscomesola/testi/graminacee.html>
informazione tecniche sulle graminacee.
- <http://www.arcgis.com/>
piattaforma per gestire mappe geografiche.
- <http://www.youtube.com/watch?v=FvVHJP9HhJw>
Video di ARPA Umbria su come utilizzare il captaspore.