Concentrazione del particolato atmosferico (Particulate matter,PM10) nell’aria della città di Taranto in relazione alle condizioni ambientali

Giorgia Desantis, Alessandro Mori, Eros Quésada,Lorenzo Panunzio,Cristiana Poggi



Caption

### Indice

1)Introduzione 1.1 Il particolato atmosferico 1.2 Origine 1.3 Effetti sulla salute umana 1.4 Legislazione 1.5 Obiettivi della ricerca 2) Materiali e metodi 2.1 I dati 2.2 Analisi esplorativa 2.2.1 Diagrammi,Box plot,Grafici di dispersione a coppie 2.2.2 Analisi delle componenti principali (ACP o PCA) 2.3 Analisi statistica 2.3.1 La regressione multipla 2.3.2 GAM (Generalised additive modelling) 3) Conclusione 4) Bibliografia

### 1) Introduzione

Negli ultimi decenni diversi studi di stampo epidemiologico e tossicologico hanno indicato l’inquinamento atmosferico come responsabile di effetti avversi sulla salute dell’uomo e dell’ambiente, (APAT, 2007). Le evidenze scientifiche hanno documentato un’ampia gamma di esiti sanitari e ambientali negativi dovuti alla presenza di particolato atmosferico (Particulate matter, PM) nell’aria. La presenza di particolato atmosferico nell’aria è infatti considerata la causa di irritazioni dell’apparato respiratorio fino al tumore al polmone, e la concausa dei cambiamenti climatici (Quarol et al., 2001).

#### 1.1) Il particolato atmosferico

Le particelle costituenti il PM variano per composizione, origine e dimensione, (APAT, 2007). Il particolato atmosferico è una miscela complessa di particelle (carbonio, ammonio, nitrati, solfiti, polveri minerali, elementi in traccia e acqua), (Hueglin et al., 2005), le cui sostanze possono presentarsi miscelate allo stato condensato (solido o liquido).  
Il particolato atmosferico viene notoriamente classificato in base al diametro aerodinamico delle particelle che lo costituiscono, poiché da questa caratteristica dipende il tempo di sospensione e quindi la permanenza nell’aria, (APAT, 2007):

• PM10 : particelle di diametro aereodinamico inferiore ai 10 micron, raggiungono la parte sommitale del sistema respiratorio e, parzialmente, i polmoni. Le precipitazioni rimuovono queste particelle dall’aria dopo poche ore dalla loro emissione. • PM2.5: particelle di diametro aereodinamico inferiore ai 2.5 micron, sono le più pericolose poiché penetrano a fondo nei polmoni e vi permangono. Le precipitazioni non rimuovono considerevolmente dall’aria queste particelle, così che esse possono permanere in atmosfera per giorni o addirittura settimane dopo la loro emissione.

#### 1.2) Origini del particolato atmosferico

Le origini del particolato atmosferico sono varie e se ne riconoscono di antropogeniche e di naturali. Tra i processi antropici che emettono particolato atmosferico nell’aria ricordiamo: la combustione di carburante per veicoli (sia diesel che benzina), i combustibili solidi (carbone, lignite e biomassa), il riscaldamento domestico, le attività industriali (come costruzioni, miniere, manifattura di cemento, ceramica e mattoni, fusione), l’erosione del manto stradale a causa del traffico, le abrasioni di freni e pneumatici, i lavori in grotte e miniere (APAT, 2007).

#### 1.3) Effetti sulla salute umana

Gli effetti dell’inquinamento atmosferico da PM10 sulla salute umana sono stati oggetto di intensi studi negli ultimi anni, (Hueglin et al., 2005). L’esposizione a particolato atmosferico (PM10) è stata associata ad aumenti della mortalità e dei ricoveri ospedalieri dovuti a malattie respiratorie e cardiovascolari. Gli effetti sono stati osservati per esposizioni a livelli molto bassi di concentrazione di PM10 e non è chiaro se esista una concentrazione soglia per il particolato al di sotto della quale non sono probabili effetti sulla salute. Studi in vitro ed in vivo su animali ed esseri umani hanno rivelato potenti effetti pro – infiammatori che coinvolgono cellule epiteliali polmonari. Il particolato atmosferico una volta nel polmone viene assorbito dalle cellule epiteliali ed attaccato dai macrofagi. L’attacco dei macrofagi alveolari instaura la secrezione da parte delle cellule del polmone di citochinine e mediatori immunitari nel sangue che stimolano l’attivazione dei globuli bianchi (neutrofili e linfociti), (Brunekreef e Holgate, 2002). Nel breve periodo l’infiammazione porta ad un danno tissutale acuto conosciuto come metaplasia polmonare (Terzano, 2006), rilevabile per la presenza di proteine come il recettore del fattore di crescita epidermico (EGFR) che evidenziano tentativi di riparazione dell’organo, e produzione continua di muco che porta ad un restringimento del lume polmonare. Attraverso l’attivazione delle vie di segnalazione dello stress dall’epitelio ai microvasi polmonari, vengono generati fattori che influenzano la coagulazione del sangue il che può condurre a disturbi circolatori come l’aritmia, (Brunekreef e Holgate, 2002). Nel ’98 l’OMS ha indicato le elevate concentrazioni di PM10 in otto città italiane (media pesata 52.6 g/m3), come causa di 3500 decessi, (APAT, 2007).

#### 1.4 Legislazione

In numerosi Paesi, le concentrazioni del particolato atmosferico di entrambe le categorie dimensionali vengono misurate regolarmente da decenni. In Italia, con poche eccezioni (Firenze, ad esempio), solo il PM10 è misurato routinariamente in centraline fisse di monitoraggio, poiché esso è considerato una buona misura del complesso mix di inquinanti solidi e gassosi creati dal trasporto, dalla combustione di carburanti per veicoli e dalla produzione di energia elettrica e termica, (APAT, 2007; Gasparinetti, 2015).

Alla luce delle conseguenze sanitarie dovute a concentrazioni superiori alla concentrazione considerata naturale per il PM10 (6 g/m3), su indicazione dell’Organizzazione mondiale della sanità (OMS), il Consiglio Europeo ha emanato la Direttiva 1999/30/EC (Unione Europea, 1999) nella quale vengono introdotti due differenti limiti per il PM10 per la protezione della salute umana: un limite per la media giornaliera di 50 g/m3, da non superarsi per più di sette volte in un anno, e uno per quella annuale, 40 g/m3, (Lena e Pirrollo, 2010). Questi limiti sono stati successivamente confermati nella Direttiva 2008/50/CE (Gazzetta Ufficiale dell’Unione Europea, 2008)

#### 1.5 Obiettivi della ricerca.

Numerosi studi hanno indagato il potenziale ruolo dei fattori meteorologici, nella diminuzione della concentrazione del particolato atmosferico in alcune città e zone rurali europee (i.e: Sanchez – Reina et al., 2006; Olszowsky, 2015;), con esiti positivi: la concentrazione di particolato atmosferico con diametro aereodinamico 10g (PM10) viene influenzata dalla presenza di precipitazioni atmosferiche e varia a seconda della direzione del vento. Nella presente ricerca viene esaminata la concentrazione di PM10 in una delle città più inquinate d’Italia, Taranto. L’obiettivo della ricerca è quello di indagare la variazione delle concentrazioni di PM10 e comprendere la misura in cui le variabili ambientali ne regolino la concentrazione.

### 2)Materiali e metodi

#### 2.1) I dati

Sono stati analizzati i valori giornalieri (espressi in media giornaliera) di concentrazione della polevere sottile PM10. Il data frame è composto da 365 osservazioni su di 11 variabili. Le variabili sono: Media(valore medio giornaliero), Mediana (valore mediano giornaliero), Massima (valore massimo giornaliero), Temperatura, Umidità relativa, Pioggia, Radiazione solare, Pressione, Velocità del vento, Direzione del vento (secondo la rosa dei venti), Data del rilevamento.

#### 2.2)Analisi esplorativa

Per visualizzare graficamente il comportamento dei dati è stata condotta un’analisi esplorativa degli stessi.Sono stati disegnati dei diagrammi, dei grafici di dispersione a coppie, dei boxplot ed è stata condotta l’analisi delle componenti principali (ACP o PCA).

#setto la WD ----  
pm10 <- read.csv2("data/pm2010.csv", dec = ".")  
  
#questo è un pacchetto di pacchetti. Contiene molti pacchetti utili :-)  
  
library(tidyr)  
#creO variabile per anno, giorno, mese  
pm10\_a <- separate(pm10, col = "data", into = c("anno","mese","giorno"), sep = "-")  
  
pm10\_a$mese2 <- as.numeric(pm10\_a$mese)  
  
  
pm10\_a$mese2

## [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1  
## [24] 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2  
## [47] 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3  
## [70] 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 4 4  
## [93] 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4  
## [116] 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5  
## [139] 5 5 5 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6 6  
## [162] 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7  
## [185] 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7  
## [208] 7 7 7 7 7 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8 8  
## [231] 8 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9  
## [254] 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 10 10 10  
## [277] 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10  
## [300] 10 10 10 10 10 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11  
## [323] 11 11 11 11 11 11 11 11 11 11 11 11 12 12 12 12 12 12 12 12 12 12 12  
## [346] 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12 12

pm10\_a$giorno2 <- as.numeric(pm10\_a$giorno)  
  
# stagioni: primavera (21-03 / 20-06), estate (21-06/22-09), autunno (23-09 /21-12)  
# inverno(22-12/20-03)  
  
#divido in stagioni ----  
getSeason <- function(DATES) {  
 WS <- as.Date("2010-12-22", format = "%Y-%m-%d") # Winter Solstice  
 SE <- as.Date("2010-3-21", format = "%Y-%m-%d") # Spring Equinox  
 SS <- as.Date("2010-6-21", format = "%Y-%m-%d") # Summer Solstice  
 FE <- as.Date("2010-9-23", format = "%Y-%m-%d") # Fall Equinox  
 d <- as.Date(strftime(DATES, format="2010-%m-%d"))  
   
 ifelse (d >= WS | d < SE, "Winter",  
 ifelse (d >= SE & d < SS, "Spring",  
 ifelse (d >= SS & d < FE, "Summer", "Fall")))  
}  
  
  
k <- getSeason(pm10$data)  
  
  
  
pm10\_a$Stagione<- k

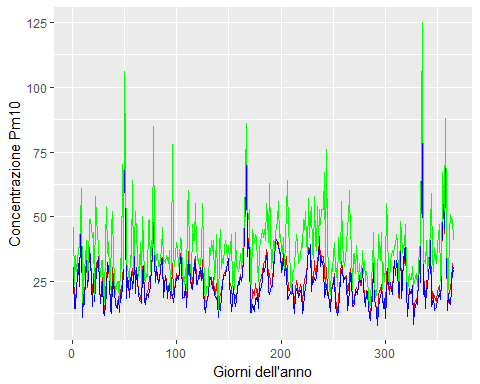
data\_hills<-(pm10\_a[1:10])  
data\_hills$stagione<-pm10\_a$Stagione

#un modo carino per selezionare i colori  
color = grDevices::colors()[grep('gr(a|e)y', grDevices::colors(), invert = T)]  
  
# questo comando crea un oggetto che si chiama color che contiene tutti e 433 i colori base di R escluse le scale di grigio  
# poi basta fare col = sample(color, n) all'intero di una funzione che usa come argument col  
# dove n è il numero di colori che devi utilizzare  
  
#alternativa se 433 colori sono troppi  
library(RColorBrewer)  
qual\_col\_pals = brewer.pal.info[brewer.pal.info$category == 'qual',]  
col\_vector = unlist(mapply(brewer.pal, qual\_col\_pals$maxcolors, rownames(qual\_col\_pals)))  
# questa funzione produce un vettore di 74 colori

##### 2.2.1) Diagrammi, Boxplot, Grafici di dispersione a coppie

Il diagramma o serie storica delle concentrazioni di Pm10 (Fig 1) è stato disegnato mantenendo in ascissa il tempo espresso in giorni ed in ordinata le concentrazioni medie, massime e mediane di pm10 .Questo ha mostrato un’andamento periodico per le tre variabili che,per media e mediana risulta essere il medesimo, mentre per la massima presenta dei picchi di maggiore intensità rispetto a le due.

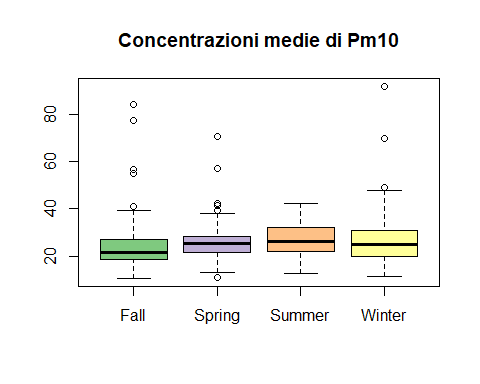
library(ggplot2)  
pm10$data2 <- seq(1,365,1)  
  
  
ggplot(pm10, aes(x = data2))+  
 geom\_path(aes(y = media), color = "red")+  
 geom\_path(aes(y = mediana), color = "blue")+  
 geom\_path(aes(y = max), color = "green")+  
 xlab("Giorni dell'anno") + ylab("Concentrazione Pm10")

 ###### Fig 1)Serie temporale concentrazione media,mediana e massima di Pm10 durante l’anno 2009(espresso in giorni)

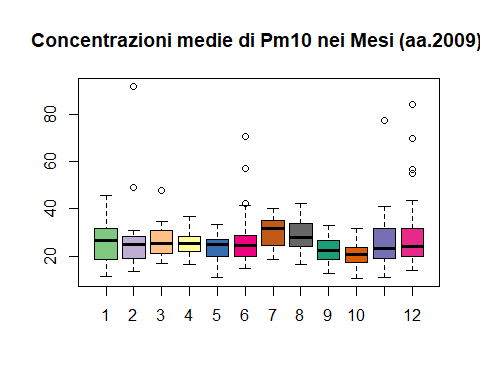
##### Boxplot

I Boxplot, “grafici a baffi e scatole”, sono stati utili per mostrare in modo compatto la distribuzione delle variabili ed hanno permesso di individuare la presenza di possibili valori anomali (outliers) .Questi sono il disegno su di un piano cartesiano di un rettangolo, i cui estremi rappresentano il primo e terzo quartile (Q1 e Q3) e sono tagliati a metà da una linea che rappresenta la mediana (Q2). Il minimo della distribuzione viene indicato con (Q0), mentre il massimo con (Q4) i quali corrispondono agli estremi delle due righe (baﬃ).I baffi si trovano per convenzione ad una distanza di 1.5 volte la distanza interquartile (Q3-Q1) a partire rispettivamente dal primo dal terzo quartile.I valori anomali sono rappresentati con dei punti al di fuori dell’intervallo compreso tra i baffi.

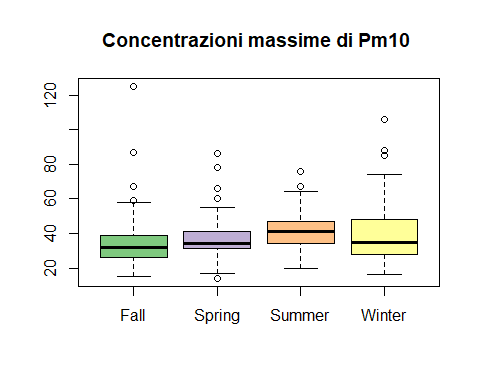
boxplot(data\_hills$media~data\_hills$stagione,col=col\_vector,main="Concentrazioni medie di Pm10")

 ###### Fig 2)Boxplot delle concentrazioni medie di Pm10 durante le stagioni nell’anno 2009 a Taranto.

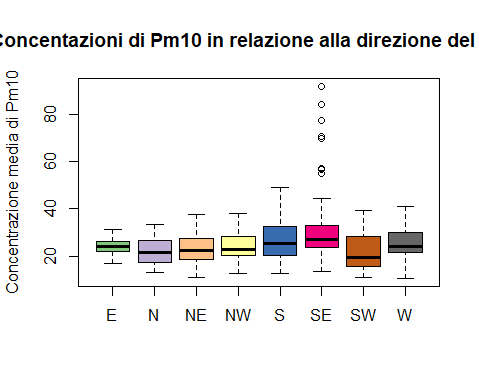
boxplot(pm10\_a$media~pm10\_a$mese2,col=col\_vector,main="Concentrazioni medie di Pm10 nei Mesi (aa.2009) ")

 ###### Fig 2.1)Boxplot delle concentrazioni medie di Pm10 durante i mesi dell’anno 2009 a Taranto

boxplot(data\_hills$max~data\_hills$stagione,col=col\_vector, main="Concentrazioni massime di Pm10 ")

 ###### Fig2.2)Boxplot delle concentrazioni massime di Pm10 durante le stagioni dell’ anno 2009 a Taranto

plot(data\_hills$dv,data\_hills$media, main=" Concentazioni di Pm10 in relazione alla direzione del vento",ylab="Concentrazione media di Pm10", col=col\_vector)

 ###### Fig2.3)Boxplot delle concentrazioni medie di Pm10 in relazione alla direzione del vento dell’anno 2009 a Taranto

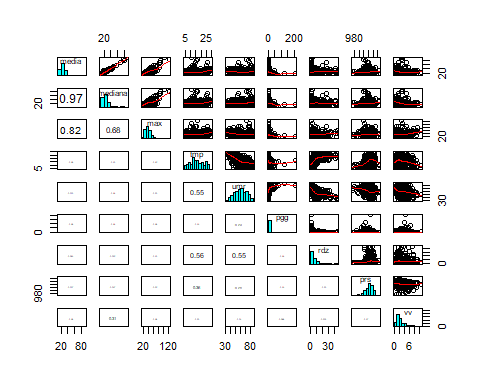
Sono stati disegnati dei boxplot relativi alle concentrazioni medie (Fig 2) e massime(Fig 2.2) di Pm10 in relazione alle stagioni per verificare se vi fosse presenza di valori anomali o variazioni evidenti in una determinata stagione rispetto alle altre e per valutarne l’andamento generale. Dall’osservazione di questi è emerso un andamento sinusoidale maggiormente visibile nei boxplot relativi al mese (Fig2.1) che indica la forte presenza di stagionalità nelle concentrazioni di Pm10. Dall’osservazione dei Boxplot sulla concentrazione media di Pm10 in relazione alla direzione del vento (secondo la rosa dei venti),(Fig 2.3) è invece emersa la forte influenza del vento proveniente da SudEst, la quale presenta numerosi valori anomali, e quindi concetrazioni di Pm10 più alte rispetto alle altre direzioni.

##### Grafico di dispersione a coppie

Il grafico a coppie o matrice di dispersione (scatter plot matrice), è la rappresentazione grafica per lo studio delle rappresentazioni numeriche bivariate. Grazie al grafico abbiamo una visualizzazione complessiva delle variabili ambientali. Nel grafico avremo in diagonale gli isogrammi, al di sotto dei quali abbiamo il coefficiente di correlazione di Pearson (r(Xi,xj)=Cov(X,Xj)/Var(Xi)Var(Xj)), e al di sopra i grafici a dispersione con le curve di regressione in rosso.

panel.hist <- function(x, ...)  
{  
 usr <- par("usr"); on.exit(par(usr))  
 par(usr = c(usr[1:2], 0, 1.5) )  
 h <- hist(x, plot = FALSE)  
 breaks <- h$breaks; nB <- length(breaks)  
 y <- h$counts; y <- y/max(y)  
 rect(breaks[-nB], 0, breaks[-1], y, col = "cyan", ...)  
}  
  
  
  
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...)  
{  
 usr <- par("usr"); on.exit(par(usr))  
 par(usr = c(0, 1, 0, 1))  
 r <- abs(cor(x, y))  
 txt <- format(c(r, 0.123456789), digits = digits)[1]  
 txt <- paste0(prefix, txt)  
 if(missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)  
 text(0.5, 0.5, txt, cex = cex.cor \* r)  
}

data\_pca <- pm10\_a[,c(1:9)]  
pairs(data\_pca,upper.panel =panel.smooth,lower.panel=panel.cor,diag.panel = panel.hist)

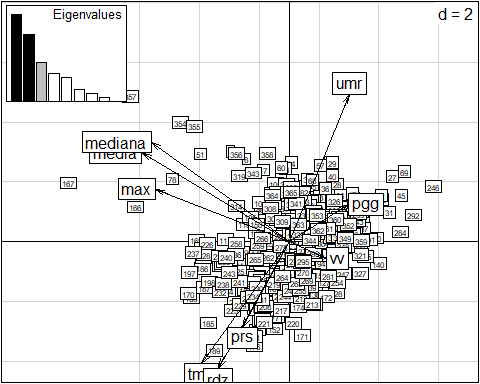
 ###### Fig 2.4) Grafici a dispersione, istogrammi e correlazioni, in rosso curve di rigressione tra coppie di variabili

##### Analisi delle componenti principali (ACP o PCA)

### PCA ANALYSIS ----  
require(ade4)

## Loading required package: ade4

#estraggo da pm10 solo le variabili numeriche da inserire nella PCA escludendo giorno mese e anno che sono da considerarsi factors  
data\_pca <- pm10\_a[,c(1:9)]  
  
# con nf = 3 gli stò dicendo di tenere tre assi  
pcapm <- dudi.pca(data\_pca, scannf = FALSE, nf = 3)  
  
#faccio il biplot, cos'è clab.row?  
# clab.row serve a gestire la grandezza dei quadratini delle osservazioni  
# clab.col invece controlla le variabili  
scatter(pcapm, clab.row = 0.5)

 ###### Fig 2.5) Biplot delll’analisi in componenti principali basata sulla matrice di correlazione.

library(broom)  
pcapm$c1

## CS1 CS2 CS3  
## media -0.5148158 0.31181443 0.02609813  
## mediana -0.4812498 0.34765592 -0.10785132  
## max -0.4683100 0.18372295 0.37203784  
## tmp -0.3058405 -0.42397086 -0.17915070  
## umr 0.2086416 0.51464826 -0.29406707  
## pgg 0.2067228 0.12934962 -0.03608484  
## rdz -0.2474513 -0.44087121 0.01414590  
## prs -0.1624832 -0.29843062 -0.31322712  
## vv 0.1288881 -0.05813173 0.79441234

###### Tabella 1) Punteggi delle variabili sui tre assi principali

#calcolo gli autovalori  
pcapm$eig / sum(pcapm$eig)

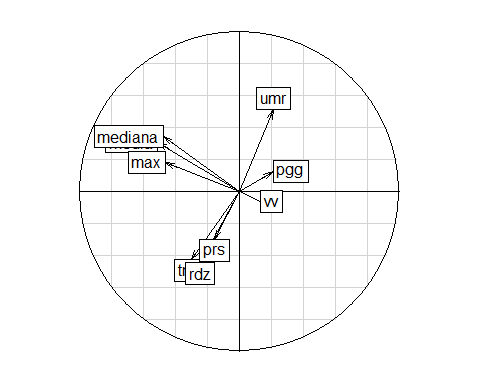
## [1] 0.3219695419 0.2473643668 0.1446821326 0.1041546381 0.0881580394  
## [6] 0.0455184307 0.0308126012 0.0168601788 0.0004800706

#calcolo la somma cumilata degli autovalori  
cumsum(pcapm$eig / sum(pcapm$eig)) # con i primi tre assi spiego circa il 70% che non è male

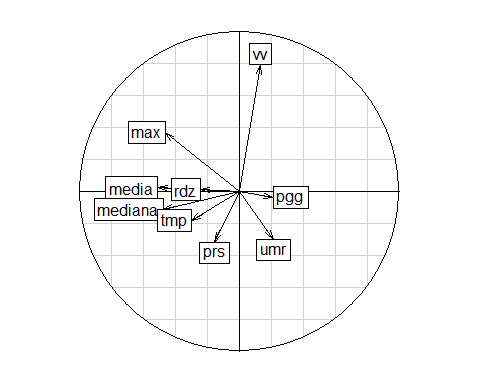
## [1] 0.3219695 0.5693339 0.7140160 0.8181707 0.9063287 0.9518471 0.9826598  
## [8] 0.9995199 1.0000000

Attraverso l’analisi delle componenti principali(ACP o PCA,Principal component analysis) è stato possibile rappresentare l’insieme dei dati in uno spazio di dimensione ridotta, definendo le direzioni di massima informazione. L’analisi in componenti principali (PCA) ha come scopo primario la riduzione di un numero più o meno elevato di variabili (rappresentanti le caratteristiche del fenomeno analizzato) in alcune variabili latenti, dette componenti principali. Queste identificano i pattern di variabilità rilevanti nell’insieme dei dati disponibili. Solitamente, infatti, si ha a che fare con insiemi di dati altamente multivariati; per poter rappresentare graficamente questi dati conservando il massimo dell’informazione si rende necessario riportare le variabili in uno spazio di 2 o 3 dimensioni, in questo caso sono stati scelti 3 assi. Nell’analisi in componenti principali si è preferito tenere in considerazione le sole variabili quantitative, per poter visualizzare in maniera chiara la distribuzione dell’informazione presente nel dataset. Dall’ osservazione del biplot (Fig 2.5) e la lettura dei punteggi (Tabella 1) è stato possibile individuare il contributo di ciascuna variabile agli assi. Attraverso la somma cumulata degli autovalori, è stato possibile calcolare la percentuale di informazione spiegata dalle componenti.Le 3 componenti principali spiegano il 71% dell’ informazione presente nel dataset; sono stati poi caratterizzati gli assi ( Fig 2.6-2.7-2.8).Grazie al biplot ed ai punteggi comprendiamo quali variabili danno maggior contributo agli assi: le concentrazioni di Pm10 (massime, medie e mediane) per il primo asse, le variabili ambientali temperatura, pressione e radiazione solare con correlazione negativa rispetto all’umidità, per il secondo asse (la pioggia sembra essere al litime con un punteggio abbastanza basso) la velocità del vento per il terzo asse.

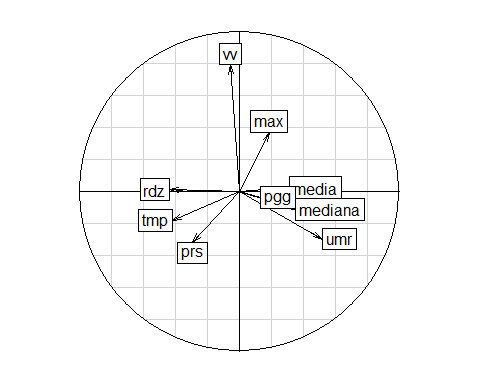
#cerchio di correlazione, i punteggi delle varibili sono plottati su una circonferenza di raggio unitario  
s.corcircle(pcapm$c1, xax = 1, yax = 2) #plotto prima e seconda

 ###### Fig 2.6)Cerchio di correlazione primo e secondo asse principali

s.corcircle(pcapm$c1, xax = 1, yax = 3) #plotto prima e terza

 ###### Fig 2.7) Cerchio di correlazione primo e terzo asse principali

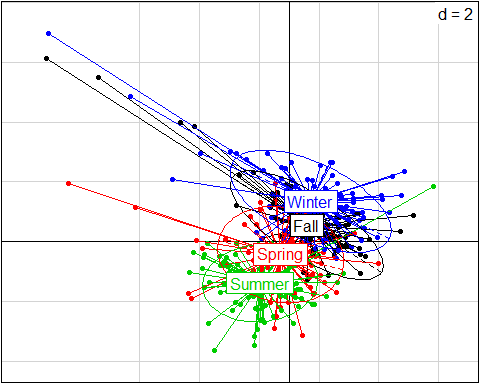
s.corcircle(pcapm$c1, xax = 2, yax = 3) #plotto seconda e terza

 ###### Fig 2.8) Cerchio di correlazione secondo e terzo asse principali

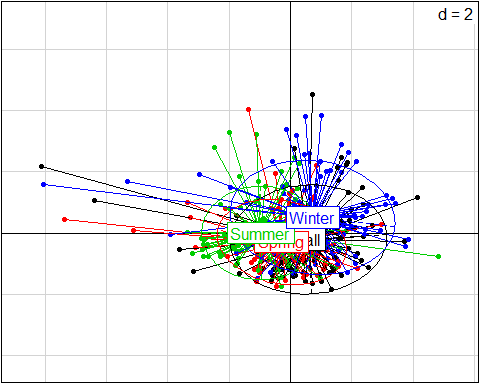
##### Cerichi di correlazione tra gli assi e le variabili qualitative

Per comprendere se le variabili qualitative omesse dalla PCA avessero qualche tipo di influenza sulle componenti principali e nella caratterizzazione degli assi sono stati disegnati dei cerchi di correlazione tra variabili qualitative (Stagione e Direzione del vento) e le componenti principali. Nei cerchi di correlazione (Fig 2.10- 2.11-2.12) relativi alle stagioni, le ellissi si dispongono trasversalmente e si nota una maggiore variazione sulla direzione della seconda componente rispetto alla prima. Questo significa che c’è una certa variazione dei fattori ambientali durante l’anno: temperatura , radiazione solare e pressione elevate in estate e invece umidità e precipitazioni elevate in inverno. In termini di concentrazione di pm10 c’è una variazione ma sembrerebbe essere inferiore rispetto a quelle delle variabili ambientali. Tuttavia, l’asse maggiore delle ellissi per autunno e inverno è lungo la direzione della prima componente il che indica che in quei mesi c’è una maggiore varibilità ma in termini di media stagionale sono identici. La primavera è praticamente un cerchio il cui centro combacia con il centro del piano fattoriale indicando che le condizioni corrispondono alla media del sistema. Se ne deduce che il fattore stagionale in termini di concentrazione di pm10 e fattori ambientali è piuttosto evidente.

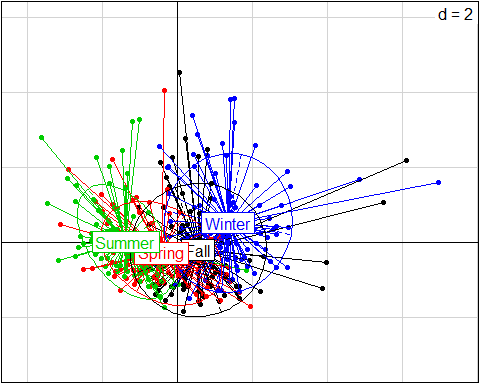
s.class(pcapm$li, factor(pm10\_a$Stagione), xax = 1, yax = 2, col = c(1,2,3,4))

 ###### Fig 2.9)Cerchio di correlazione primo e secondo asse in relazione alla stagione

s.class(pcapm$li, factor(pm10\_a$Stagione), xax = 1, yax = 3, col = c(1,2,3,4))

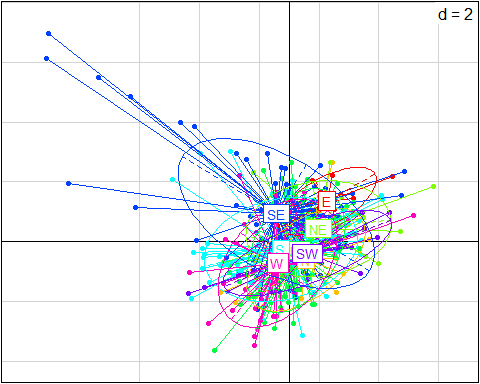
 ###### Fig 2.10) Cerchio di correlazione primo e terzo asse in relazione alla stagione

s.class(pcapm$li, factor(pm10\_a$Stagione), xax = 2, yax = 3, col = c(1,2,3,4))

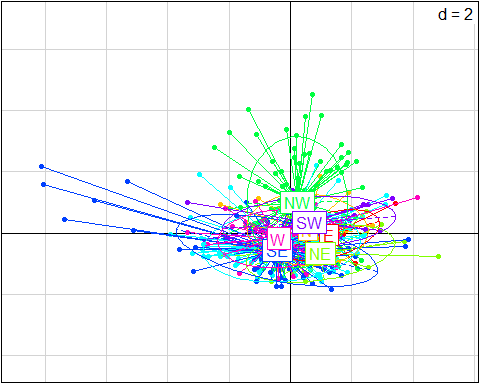
 ###### Fig 2.11) Cerchio di correlazione secondo e terzo asse in relazione alla stagione

Per quanto rigarda il fattore direzione del vento (espresso secondo la rosa dei venti) i cerchi di correlazione hanno mostrato un’influenza maggiore solo in inverno con una corrispondente concentrazione di Pm10 più bassa, cosa plausibile se si pensa che in inverno il vento forte tende a disperdere il Pm 10. I venti da SE mostrano una variabilità estrema sia in termini di concentrazione di pm10 sia per quanto riguarda gli altri fattori ambientali, mentre i venti da NW sono i più intensi, ma non sembrano corrispondere ad una riduzione nella concentrazione di Pm10.

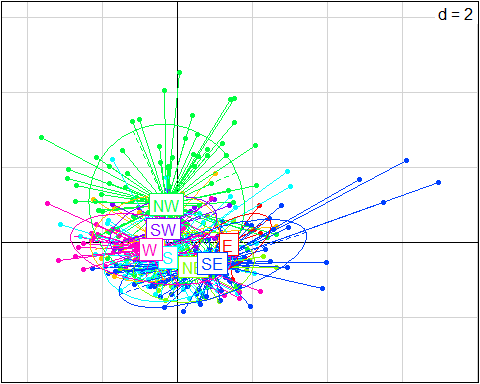
#DIREZIONE DEL VENTO ----  
s.class(pcapm$li, factor(pm10\_a$dv), xax = 1, yax = 2, col = rainbow(nlevels(pm10$dv)),clabel = .8)

 ###### Fig 2.13)Cerchio correlazione primo e secondo asse in relazione alla direzione del vento

s.class(pcapm$li, factor(pm10\_a$dv), xax = 1, yax = 3, col = rainbow(nlevels(pm10$dv)))

 ###### Fig 2.14) Cerchio di correlazione primo e terzo asse in relazione alla direzione del vento

s.class(pcapm$li, factor(pm10\_a$dv), xax = 2, yax = 3, col = rainbow(nlevels(pm10$dv)))

 ###### Fig 2.15) Cerchio di correlazione secondo e terzo asse in relazione alla direzione del vento

#### 2.3) Analisi statistica

##### 2.3.1) La regressione lineare multipla

Sono stati stimati dei modelli di regressione lineare multipla (Tabella 2) al fine di comprendere se vi fosse una relazione lineare tra la variabile indipendente ovvero la concentrazione media di Pm10 e le variabili dipendenti ovvero le variabili ambientali. La formula del modello è: Y=o +1 x1 + 2 x2 + …p xp +i dove: -Y rappresenta la variabile indipendente (conc.media Pm10) -o rappresenta l’interecetta del modello -p xp rappresenta il coefficiente di regressione parziale che da l’intensità e la direzione della relazione tra x ed y. -i rappresenta i residui del modello che dovranno rispettare le regole di indipendenza, omoschedasticità e normalità(Distribuzione gaussiana). Il test di ipotesi alla base del modello è: Ho: i uguale a 0 H1: i diverso da 0 bove i rappresenta il coefficiente della variabile i-esima La statistica test utilizzata dal modello di regressione lineare multipla è il test t. (equazione) Al fine di evitare la multicollinearità (overfitt) dovuta all’eccessiva correlazione delle variabili media e mediana è stato scelto di escludere quest’ultima dalla stima del modello. Per identificare quale modello rappresentasse meglio l’insieme dei dati si è deciso di procedere utilizzando la procedura automatica Stepwise (Stepwise regression). Una volta stimato il modello di regressione ottimale (Tabella 2,modello2) è stata applicata la procedura che mediante una misura di adattamento (Googness of fit) ha stimato il miglior modello (Tebella 2, modello3). La procedura Stepwise si basa sull’AIC (akaike information criterion), ovvero la misura di entropia dei dati, questo è un criterio di verosimiglianza ed assume che i dati siano gaussiani.

Yi~ N(o + 1 x1 … + p xp) P(y<=y1)

Dall’output del modello scelto(tabella 3),risultano essere maggiormente significative le variabili massima, velocità del vento, ed umidità mentre leggermente significative la temperatura e la stagione Winter.Dal valore dell’r2(0.83) risulta essere un buon modello per spiegare i dati. L’analisi dei residui(Fig 2.16) conferma la bontà del modello:la linearità e l’indipendenza è spiegata graficamente attraverso il grafico dei resuidi sui valori previsti (ResidualvsFitted) in cui la nuvola dei punti è distribuita abbastanza uniformemente intorno alla linea dello zero.La normalità è verificata attraverso il Q-Q plot, in cui i punti si dispongono lungo la bisettrice. Attraverso il grafico dei residui standardizzati verso i valori stimati del modello(Scale-Location) è confermata l’omoschedasticità dei residui.Il grafico dell’autocovarianza e autocorrelazione (fig 2.18), presenta in ordinata delle linee di autocorrelazione ed in ascissa il lag (intervallo di tempo tra le osservazioni), mentre le linee tratteggiate orizzontali indicano il limite di significatività statistica. Questo risulta essere buono non mostrando valori significativi nei residui. Inoltre i residui sono stati disegnati in funzione della stagione e della direzione del vento (Fig.2.19-2.20).

mod2<-lm(media~tmp+max+vv+dv+rdz+pgg+umr+prs+stagione,data=data\_hills)  
summary(mod2)

##   
## Call:  
## lm(formula = media ~ tmp + max + vv + dv + rdz + pgg + umr +   
## prs + stagione, data = data\_hills)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.4301 -1.7859 -0.0942 1.9276 17.2747   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 24.63988 30.74508 0.801 0.42343   
## tmp 0.17805 0.07605 2.341 0.01979 \*   
## max 0.65088 0.01764 36.888 < 2e-16 \*\*\*  
## vv -1.15876 0.14812 -7.823 6.28e-14 \*\*\*  
## dvN 0.80524 1.69863 0.474 0.63576   
## dvNE 1.55747 1.64020 0.950 0.34300   
## dvNW -2.81935 1.50105 -1.878 0.06118 .   
## dvS 0.48760 1.48640 0.328 0.74308   
## dvSE 1.78856 1.48775 1.202 0.23011   
## dvSW -1.49605 1.68906 -0.886 0.37638   
## dvW -1.76805 1.54270 -1.146 0.25255   
## rdz 0.02402 0.03506 0.685 0.49377   
## pgg -0.01984 0.01620 -1.225 0.22158   
## umr 0.09628 0.03146 3.061 0.00238 \*\*   
## prs -0.02817 0.02994 -0.941 0.34726   
## stagioneSpring 0.19683 0.71281 0.276 0.78261   
## stagioneSummer -1.04027 0.95300 -1.092 0.27578   
## stagioneWinter 1.53777 0.76687 2.005 0.04571 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.059 on 347 degrees of freedom  
## Multiple R-squared: 0.8363, Adjusted R-squared: 0.8283   
## F-statistic: 104.3 on 17 and 347 DF, p-value: < 2.2e-16

b<-summary(mod2)  
AIC(mod2)

## [1] 2077.992

mod1<-lm(max~tmp+media+vv+rdz+pgg+umr+prs,data=data\_hills)  
summary(mod1)

##   
## Call:  
## lm(formula = max ~ tmp + media + vv + rdz + pgg + umr + prs,   
## data = data\_hills)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -20.080 -3.757 -0.456 2.412 36.087   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -51.63442 40.09269 -1.288 0.1986   
## tmp -0.08963 0.07050 -1.271 0.2044   
## media 1.21509 0.03310 36.707 < 2e-16 \*\*\*  
## vv 2.08902 0.18615 11.222 < 2e-16 \*\*\*  
## rdz -0.06081 0.04607 -1.320 0.1877   
## pgg -0.00687 0.02236 -0.307 0.7588   
## umr -0.22329 0.03804 -5.870 9.95e-09 \*\*\*  
## prs 0.06642 0.03929 1.691 0.0918 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5.889 on 357 degrees of freedom  
## Multiple R-squared: 0.811, Adjusted R-squared: 0.8073   
## F-statistic: 218.8 on 7 and 357 DF, p-value: < 2.2e-16

a<-summary(mod1)  
AIC(mod1)

## [1] 2340.035

mod2s=step(mod2,direction="both")

## Start: AIC=1040.17  
## media ~ tmp + max + vv + dv + rdz + pgg + umr + prs + stagione  
##   
## Df Sum of Sq RSS AIC  
## - rdz 1 7.7 5723.8 1038.7  
## - prs 1 14.6 5730.7 1039.1  
## - stagione 3 82.7 5798.8 1039.4  
## - pgg 1 24.7 5740.8 1039.7  
## <none> 5716.1 1040.2  
## - tmp 1 90.3 5806.4 1043.9  
## - umr 1 154.3 5870.4 1047.9  
## - dv 7 748.6 6464.7 1071.1  
## - vv 1 1008.1 6724.2 1097.5  
## - max 1 22414.8 28130.9 1619.8  
##   
## Step: AIC=1038.66  
## media ~ tmp + max + vv + dv + pgg + umr + prs + stagione  
##   
## Df Sum of Sq RSS AIC  
## - prs 1 16.5 5740.3 1037.7  
## - pgg 1 24.2 5748.0 1038.2  
## - stagione 3 92.0 5815.8 1038.5  
## <none> 5723.8 1038.7  
## + rdz 1 7.7 5716.1 1040.2  
## - tmp 1 108.2 5832.0 1043.5  
## - umr 1 146.9 5870.7 1045.9  
## - dv 7 749.3 6473.1 1069.6  
## - vv 1 1011.3 6735.1 1096.0  
## - max 1 22409.5 28133.3 1617.9  
##   
## Step: AIC=1037.71  
## media ~ tmp + max + vv + dv + pgg + umr + stagione  
##   
## Df Sum of Sq RSS AIC  
## - pgg 1 23.8 5764.0 1037.2  
## <none> 5740.3 1037.7  
## + prs 1 16.5 5723.8 1038.7  
## - stagione 3 115.2 5855.5 1039.0  
## + rdz 1 9.6 5730.7 1039.1  
## - tmp 1 110.4 5850.6 1042.7  
## - umr 1 160.4 5900.7 1045.8  
## - dv 7 765.9 6506.2 1069.4  
## - vv 1 1003.1 6743.4 1094.5  
## - max 1 22396.9 28137.1 1615.9  
##   
## Step: AIC=1037.22  
## media ~ tmp + max + vv + dv + umr + stagione  
##   
## Df Sum of Sq RSS AIC  
## <none> 5764.0 1037.2  
## + pgg 1 23.8 5740.3 1037.7  
## + prs 1 16.0 5748.0 1038.2  
## - stagione 3 117.7 5881.8 1038.6  
## + rdz 1 9.0 5755.0 1038.6  
## - tmp 1 104.1 5868.2 1041.8  
## - umr 1 136.9 5901.0 1043.8  
## - dv 7 763.9 6527.9 1068.6  
## - vv 1 1165.1 6929.1 1102.4  
## - max 1 22961.2 28725.3 1621.5

summary(mod2s)

##   
## Call:  
## lm(formula = media ~ tmp + max + vv + dv + umr + stagione, data = data\_hills)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -24.3552 -1.8471 -0.1752 2.1550 17.4406   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3.26439 2.98628 -1.093 0.27509   
## tmp 0.18579 0.07388 2.515 0.01236 \*   
## max 0.65279 0.01748 37.339 < 2e-16 \*\*\*  
## vv -1.17387 0.13956 -8.411 1.05e-15 \*\*\*  
## dvN 1.00496 1.67610 0.600 0.54917   
## dvNE 1.45076 1.63417 0.888 0.37528   
## dvNW -2.64756 1.47398 -1.796 0.07333 .   
## dvS 0.75688 1.46877 0.515 0.60666   
## dvSE 2.08972 1.46237 1.429 0.15390   
## dvSW -1.10744 1.67009 -0.663 0.50770   
## dvW -1.57867 1.53219 -1.030 0.30356   
## umr 0.08158 0.02829 2.884 0.00417 \*\*   
## stagioneSpring 0.26275 0.65338 0.402 0.68782   
## stagioneSummer -1.18573 0.93659 -1.266 0.20635   
## stagioneWinter 1.77057 0.73388 2.413 0.01635 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4.058 on 350 degrees of freedom  
## Multiple R-squared: 0.8349, Adjusted R-squared: 0.8283   
## F-statistic: 126.5 on 14 and 350 DF, p-value: < 2.2e-16

c<-summary(mod2s)

library(broom)  
  
tidy(mod2s)

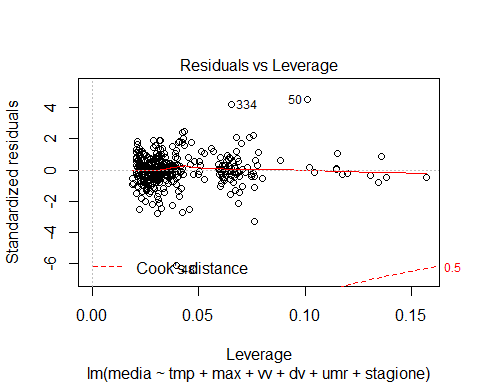
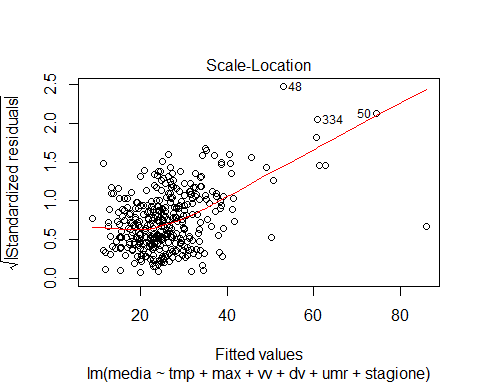
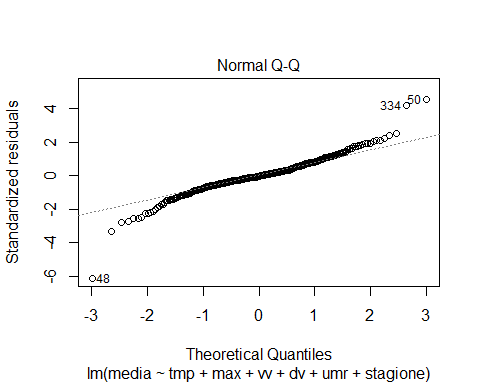
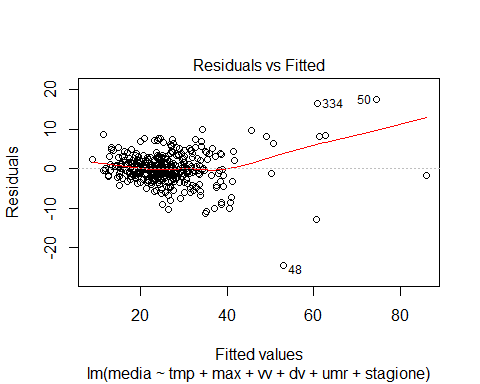
## term estimate std.error statistic p.value  
## 1 (Intercept) -3.26439265 2.98628013 -1.0931301 2.750885e-01  
## 2 tmp 0.18578848 0.07388418 2.5145909 1.236480e-02  
## 3 max 0.65278937 0.01748256 37.3394503 4.064191e-124  
## 4 vv -1.17386961 0.13956232 -8.4110782 1.045240e-15  
## 5 dvN 1.00496250 1.67609604 0.5995853 5.491706e-01  
## 6 dvNE 1.45075527 1.63417247 0.8877614 3.752788e-01  
## 7 dvNW -2.64755886 1.47397853 -1.7961991 7.332514e-02  
## 8 dvS 0.75687647 1.46876759 0.5153140 6.066590e-01  
## 9 dvSE 2.08972090 1.46237003 1.4289960 1.538973e-01  
## 10 dvSW -1.10743701 1.67008525 -0.6631021 5.077017e-01  
## 11 dvW -1.57867300 1.53218575 -1.0303405 3.035617e-01  
## 12 umr 0.08157935 0.02828999 2.8836825 4.173672e-03  
## 13 stagioneSpring 0.26275082 0.65337785 0.4021422 6.878248e-01  
## 14 stagioneSummer -1.18573213 0.93659288 -1.2660059 2.063530e-01  
## 15 stagioneWinter 1.77056876 0.73387554 2.4126281 1.635153e-02

glance(mod2s)

## r.squared adj.r.squared sigma statistic p.value df logLik  
## 1 0.8349476 0.8283455 4.058164 126.467 1.858195e-127 15 -1021.521  
## AIC BIC deviance df.residual  
## 1 2075.042 2137.44 5764.043 350

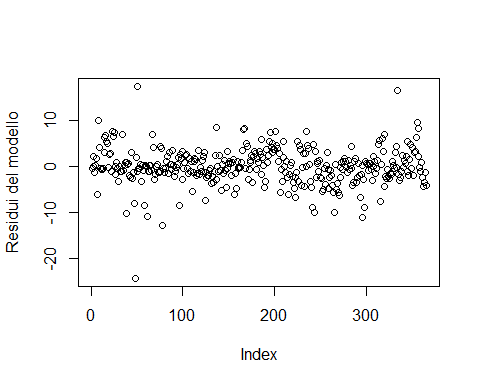
###### Tabella 3) Modello di regressione lineare multipla con procedura Stepwise.

plot(mod2s)



###### Fig 2.16) Analisi dei residui del modello

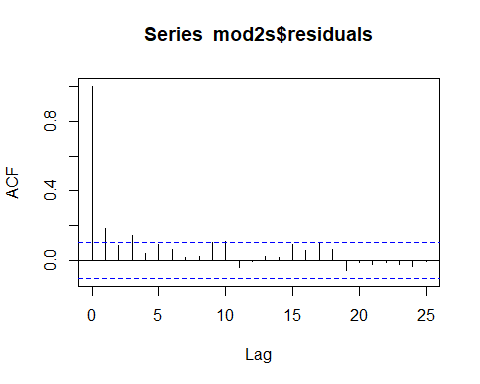
plot(mod2s$residuals,ylab="Residui del modello")



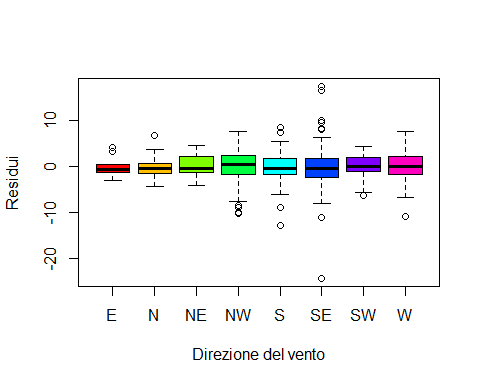
c<-summary(mod2s)

###### Fig 2.17)Analisi grafica dei residui

acf(mod2s$residuals)

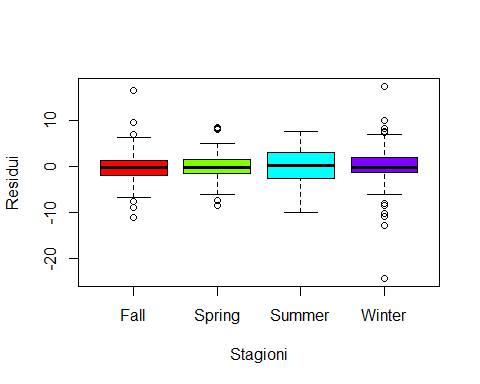
 ###### Fig 2.18) Analisi di autocorrelazione

plot(mod2s$residuals~data\_hills$dv,xlab="Direzione del vento",ylab = "Residui", col=rainbow(8))



###### Fig 2.19)Analisi dei residui in funzione della direzione del vento

data\_hills$stagione <- factor(data\_hills$stagione)  
  
plot(mod2s$residuals~data\_hills$stagione,xlab="Stagioni",ylab="Residui",col=rainbow(4))

 ######Fig 2.20) Analisi dei residui in funzione della stagione

require(xtable)

## Loading required package: xtable

tab2<-xtable(a)#modello 1  
tab3<-xtable(b)#modello 2  
tab4<-xtable(c)# modello 2s<---quello scelt  
  
AICt<-c(2340.035,2077.992,2075.042)  
variabili<-c("max~tmp+media+vv+rdz+pgg+umr+prs","media~tmp+max+vv+dv+rdz+pgg+umr+prs+stagione","media~tmp+max+vv+dv+rdz+pgg+umr+prs+stagione")  
nomemodello<-c("modello1","modello2","modello3")  
  
AICt<-as.factor(AICt)  
nomemodello<-as.factor(nomemodello)  
variabili<-as.factor(variabili)  
# non riesco a creare una fottutissima tabella in cui mettere   
# ste 4 cazzate in modo che ci sia la tabella dei modelli e quello che abbiamo scelto

# medie e norma di legge ----  
  
data\_hills$normaL<-ifelse(data\_hills$media<50.000,"low","high")  
  
data\_hills$normaL<-as.factor(data\_hills$normaL)  
  
  
#pure qui sta norma di legge l'ho dovuta accannare  
# non sono riuscita a trovare ne dati, ne crearli finti per  
# la cosa della stima ma amen ne faremo a meno

##### 2.3.2) GAM(generalised additive modelling)

I GAM sono modelli lineari generalizzati che utilizzano una distribuzione non lineare, la variabile risposta è spiegata tramite una relazione non gaussiana e la relazione (link) tra variabile risposta e variabile esplicativa può essere diversa. Il modello mette in relazione una variabile indipendente,Y con delle variabili dipendenti xi.La variabile indipendete è associata ad una famiglia di distribuzione esponenziale( distribuzione normale,binomiale o di Poisson) attraverso una funzione link g, mettendo in relazione il valore aspettato per Y con le variabili dipendenti attraverso una struttura del tipo: g(E(Y))=o +f1(x1)…+fp(xp) dove f1 è una funzione che può essere in forma parametrica (es.polinomiale) o non parametrica/ semi-parametrica o funzione di smooth.

Sono stati stimati dei GAM per comprendere se la relazione tra concentrazione media di Pm10 e le varibili ambientali potesse essere spiegata da questi. Dall’output del modello (Tabella 4) risultano essere significative le variabili giorno,umidità,temperatura e velocità del vento.Il modello presenta un r2 di 0.85 e spiega l’86.5% della devianza. Dall’analisi dei residui (Fig 2.22) risultano essere normali, indipendenti e omoschedastici.

library(mgcv)

## Loading required package: nlme

## This is mgcv 1.8-22. For overview type 'help("mgcv-package")'.

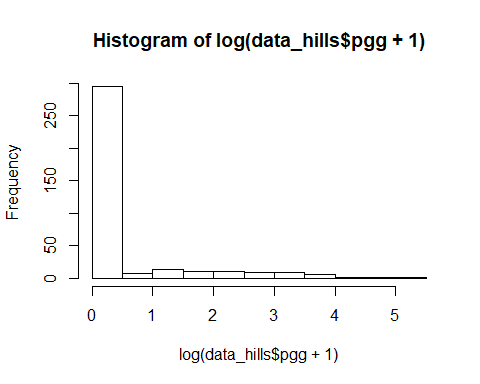
require(mgcv)  
data\_hills<-(pm10\_a[1:10])  
data\_hills$stagione<-pm10\_a$Stagione  
names(data\_hills)

## [1] "media" "mediana" "max" "tmp" "umr" "pgg"   
## [7] "rdz" "prs" "vv" "dv" "stagione"

summary(data\_hills)

## media mediana max tmp   
## Min. :10.50 Min. : 8.00 Min. : 14.00 Min. : 4.00   
## 1st Qu.:20.12 1st Qu.:18.50 1st Qu.: 30.00 1st Qu.:12.00   
## Median :24.62 Median :23.00 Median : 36.00 Median :16.00   
## Mean :26.10 Mean :24.74 Mean : 37.84 Mean :16.58   
## 3rd Qu.:30.12 3rd Qu.:28.50 3rd Qu.: 43.00 3rd Qu.:21.00   
## Max. :92.00 Max. :90.00 Max. :125.00 Max. :30.00   
##   
## umr pgg rdz prs   
## Min. :31.00 Min. : 0.000 Min. : 0.000 Min. : 975   
## 1st Qu.:51.00 1st Qu.: 0.000 1st Qu.: 2.000 1st Qu.:1007   
## Median :60.00 Median : 0.000 Median : 5.000 Median :1013   
## Mean :59.84 Mean : 3.387 Mean : 8.168 Mean :1012   
## 3rd Qu.:70.00 3rd Qu.: 0.200 3rd Qu.:11.000 3rd Qu.:1018   
## Max. :86.00 Max. :210.000 Max. :46.000 Max. :1035   
##   
## vv dv stagione   
## Min. : 0.100 NW :91 Length:365   
## 1st Qu.: 1.500 SE :82 Class :character   
## Median : 2.200 S :77 Mode :character   
## Mean : 2.836 W :46   
## 3rd Qu.: 3.700 NE :21   
## Max. :11.000 SW :20   
## (Other):28

hist(log(data\_hills$pgg+1))

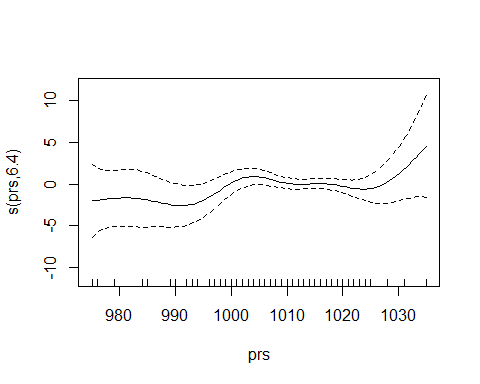
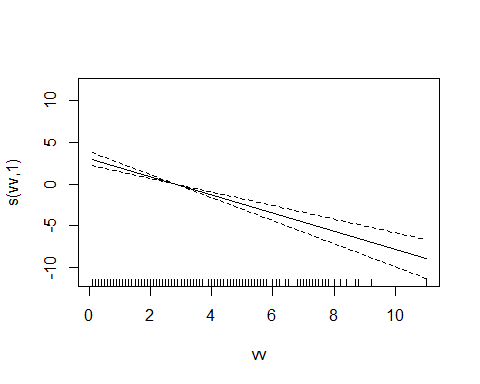
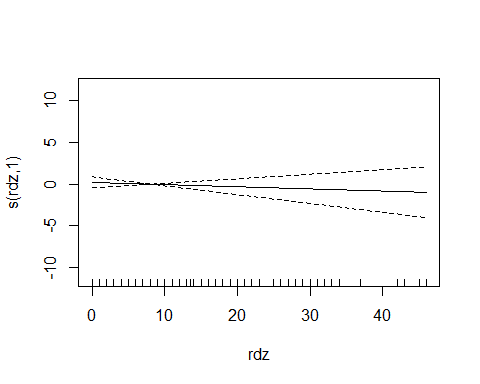
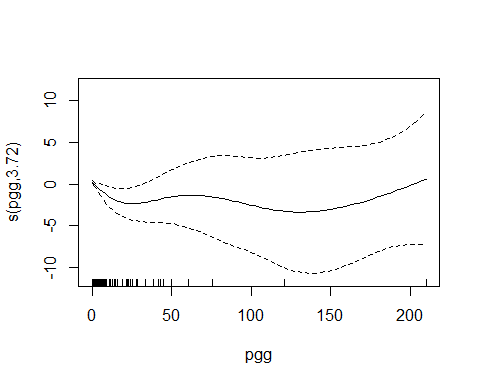
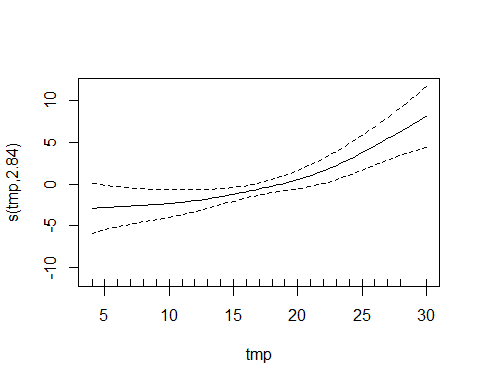
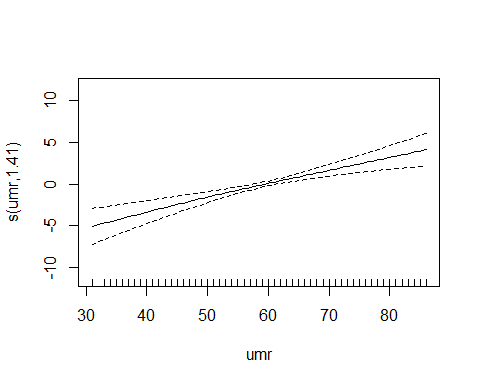
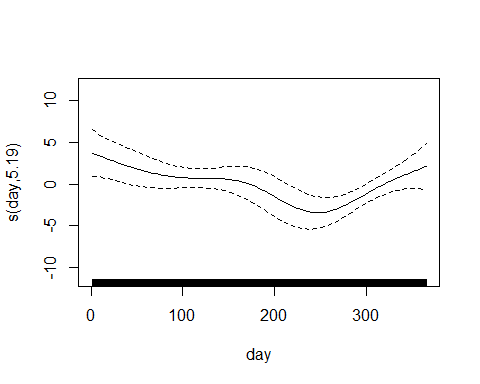


data\_hills$l.pgg <- log(data\_hills$pgg + 1)  
  
data\_hills$day <- seq(1,365,1)  
  
g<-gam(media ~ max + s(day) + s(umr) + s(tmp) + s(pgg) + s(rdz) + s(vv) + s(prs) + dv, data=data\_hills) # meglio questo  
  
library(broom)  
  
tidy(g)

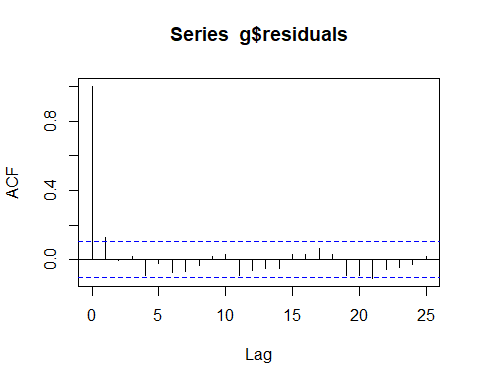
## term edf ref.df statistic p.value  
## 1 s(day) 5.192453 6.398899 3.9017872 6.868949e-04  
## 2 s(umr) 1.408334 1.717533 15.3789283 9.928873e-06  
## 3 s(tmp) 2.844137 3.589819 6.1921749 4.247636e-04  
## 4 s(pgg) 3.715445 4.432923 1.8412759 1.071066e-01  
## 5 s(rdz) 1.000000 1.000000 0.4123652 5.212114e-01  
## 6 s(vv) 1.000000 1.000000 59.2027213 1.267855e-13  
## 7 s(prs) 6.397079 7.464062 1.5766480 1.272854e-01

###### Tabella 4)Output GAM

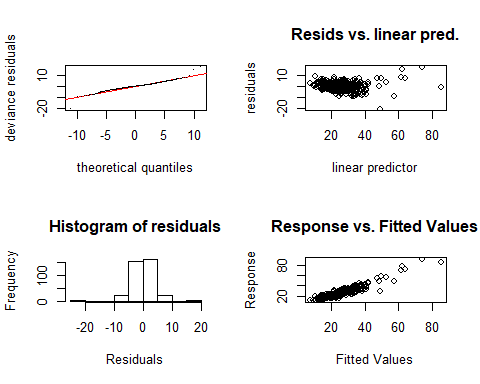
plot(g)

 ###### Fig2.21)Grafici dei GAM

acf(g$residuals)



gam.check(g)

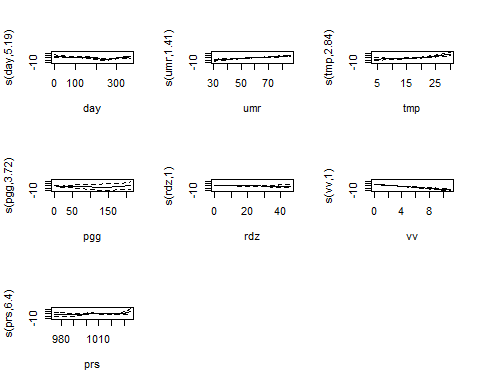


##   
## Method: GCV Optimizer: magic  
## Smoothing parameter selection converged after 17 iterations.  
## The RMS GCV score gradient at convergence was 7.077928e-07 .  
## The Hessian was positive definite.  
## Model rank = 72 / 72   
##   
## Basis dimension (k) checking results. Low p-value (k-index<1) may  
## indicate that k is too low, especially if edf is close to k'.  
##   
## k' edf k-index p-value   
## s(day) 9.00 5.19 0.87 0.005 \*\*  
## s(umr) 9.00 1.41 1.13 0.990   
## s(tmp) 9.00 2.84 0.95 0.160   
## s(pgg) 9.00 3.72 0.98 0.260   
## s(rdz) 9.00 1.00 0.95 0.165   
## s(vv) 9.00 1.00 1.04 0.745   
## s(prs) 9.00 6.40 1.03 0.705   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

g2<-predict(g,se=TRUE, type = "response")

###### Fig 2.22) Analisi dei residui del Gam

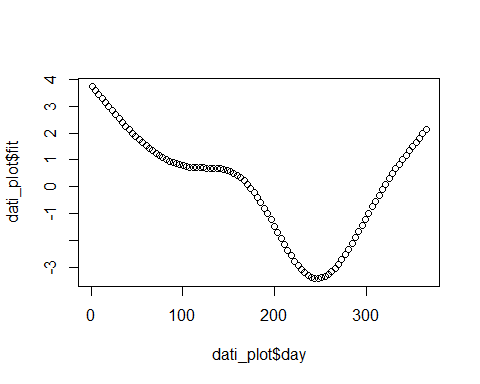
plotdata <- plot(g, pages = 1)



str(plotdata)

## List of 7  
## $ :List of 11  
## ..$ x : num [1:100] 1 4.68 8.35 12.03 15.71 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 2.81 2.68 2.56 2.47 2.39 ...  
## ..$ raw : num [1:365] 1 2 3 4 5 6 7 8 9 10 ...  
## ..$ xlab : chr "day"  
## ..$ ylab : chr "s(day,5.19)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 1 365  
## ..$ fit : num [1:100, 1] 3.74 3.59 3.43 3.28 3.13 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 31 31.6 32.1 32.7 33.2 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 2.16 2.11 2.06 2 1.95 ...  
## ..$ raw : int [1:365] 64 58 55 61 84 73 72 78 67 63 ...  
## ..$ xlab : chr "umr"  
## ..$ ylab : chr "s(umr,1.41)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 31 86  
## ..$ fit : num [1:100, 1] -5.06 -4.96 -4.85 -4.75 -4.64 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 4 4.26 4.53 4.79 5.05 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 3.04 2.94 2.84 2.75 2.67 ...  
## ..$ raw : int [1:365] 13 12 8 6 11 13 13 14 12 9 ...  
## ..$ xlab : chr "tmp"  
## ..$ ylab : chr "s(tmp,2.84)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 4 30  
## ..$ fit : num [1:100, 1] -2.9 -2.87 -2.85 -2.82 -2.79 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 0 2.12 4.24 6.36 8.48 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 0.193 0.237 0.537 0.793 0.999 ...  
## ..$ raw : num [1:365] 0 1.6 0 11.4 0.4 ...  
## ..$ xlab : chr "pgg"  
## ..$ ylab : chr "s(pgg,3.72)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 0 210  
## ..$ fit : num [1:100, 1] 0.249 -0.12 -0.485 -0.839 -1.17 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 0 0.465 0.929 1.394 1.859 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 0.656 0.619 0.582 0.544 0.507 ...  
## ..$ raw : num [1:365] 1 1 1 1 1 2 1 0 0 1 ...  
## ..$ xlab : chr "rdz"  
## ..$ ylab : chr "s(rdz,1)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 0 46  
## ..$ fit : num [1:100, 1] 0.211 0.199 0.187 0.175 0.163 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 0.1 0.21 0.32 0.43 0.54 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 0.782 0.75 0.719 0.687 0.656 ...  
## ..$ raw : num [1:365] 3.7 4.1 7.3 1.7 2.6 ...  
## ..$ xlab : chr "vv"  
## ..$ ylab : chr "s(vv,1)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 0.1 11  
## ..$ fit : num [1:100, 1] 3.01 2.89 2.77 2.64 2.52 ...  
## ..$ plot.me: logi TRUE  
## $ :List of 11  
## ..$ x : num [1:100] 975 976 976 977 977 ...  
## ..$ scale : logi TRUE  
## ..$ se : num [1:100] 4.35 4.07 3.83 3.63 3.49 ...  
## ..$ raw : int [1:365] 1001 1007 1018 1022 1010 1007 1014 1009 1002 1001 ...  
## ..$ xlab : chr "prs"  
## ..$ ylab : chr "s(prs,6.4)"  
## ..$ main : NULL  
## ..$ se.mult: num 2  
## ..$ xlim : num [1:2] 975 1035  
## ..$ fit : num [1:100, 1] -2.01 -1.95 -1.9 -1.85 -1.8 ...  
## ..$ plot.me: logi TRUE

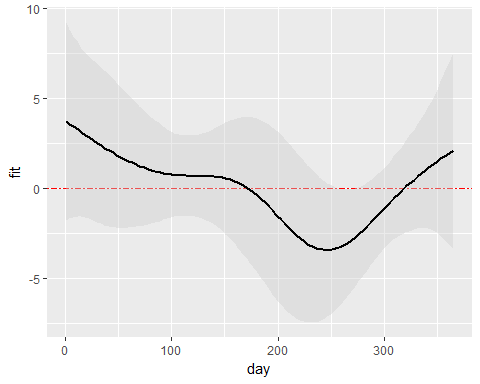
# Convert it in a dataframe  
dati\_plot <- data.frame(day = plotdata[[1]]$x, fit = plotdata[[1]]$fit,  
 se= plotdata[[1]]$se)  
  
plot(dati\_plot$day, dati\_plot$fit)



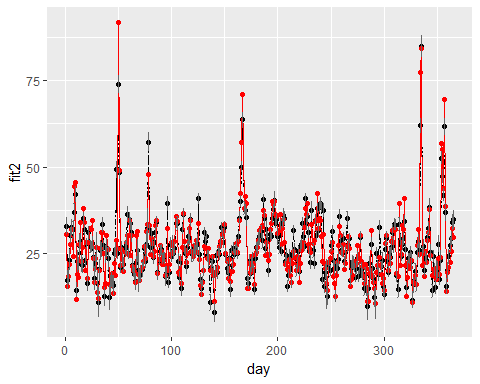
dati\_plot$day <- as.integer(dati\_plot$day)  
  
# Calculate the CI  
dati\_plot$upr <- dati\_plot$fit + (1.96 \* dati\_plot$se)  
dati\_plot$lwr <- dati\_plot$fit - (1.96 \* dati\_plot$se)  
  
data\_hills2 <- merge(dati\_plot, data\_hills, by = "day", all.y = TRUE)  
  
ggplot(data\_hills2, aes(x = day, y = fit)) +  
 geom\_hline(aes(yintercept = 0), linetype = 6, color = "red") +  
 geom\_smooth(aes(ymin = lwr, ymax = upr), fill = "gray80", color = "black", size = 1, stat = "identity")

## Warning: Ignoring unknown aesthetics: ymin, ymax

## Warning: Removed 265 rows containing missing values (geom\_smooth).



data\_hills2$fit2 <- g2$fit  
data\_hills2$se2 <- g2$se.fit  
  
data\_hills2$upr2 <- data\_hills2$fit2 + (1.96 \* data\_hills2$se2)  
data\_hills2$lwr2 <- data\_hills2$fit2 - (1.96 \* data\_hills2$se2)  
  
  
ggplot(data\_hills2, aes(x = day)) +  
 geom\_point(aes(y = fit2), color = "black") +   
 geom\_line(aes(y = fit2), color = "black", linetype = "twodash") +   
 geom\_point(aes(y = media), color = "red") +  
 geom\_line(aes(y = media), color = "red", linetype = "solid") +  
 geom\_errorbar(aes(ymin = lwr2, ymax = upr2), width = .2, color = "grey40")



# tutto questo non so come commentarlo ne ho capito cosa hai fatto...c'ho provato ma non c'ho tempo :(

### 3) Conclusione

L’obiettivo dell’analisi era quello di comprendere le relazioni che intercorrono tra la concentrazione di Pm10 e le variabili ambientali a Taranto e capire quanto le variabili ambientali ne influenzino la concentrazione. Dall’analisi è emerso che è presente un fattore stagionale nella concentrazione del pm10, in inverno la concentrazione del Pm10 presenta dei valori anomali superiori alla media dovuti probabilmente alla maggiore produzione di Pm10 a causa del riscaldamento delle abitazione e all’utilizzo delle automobili più intenso rispetto alle altre stagioni.Anche in estate vi sono dei valori anomali con concentrazioni elevate di Pm10, questo è attribuibile all’assenza delle piogge,le elevate temperature e per la presenza del vento caldo che spira da Sudest. Dalla stima del modello di regressione lineare sono risultate essere significative le variabili massima, velocità del vento, direzione del vento e in minor misura umidità e stagione invernale.Anche la stima del GAM ha confermato la significatività delle variabili velocità del vento ed umidità con l’aggiunta della temperatura.In linea generale i modelli stimati hanno quindi evidenziato l’importanza delle variabili velocità del vento, che tende a disperdere il Pm10 maggiormente e quindi in relazione a concentrazioni minori di Pm10, direzione del vento sempre in relazione alla velocità e da dove spira il vento (da mare o da terra), temperatura ed umidità che determinano minor dispersione del Pm10 e livelli più alti nelle concentrazioni.Sebbene la variabili ambientali sembrerebbero essere implicate, nella presenza di diversi livelli di concentrazione di Pm10 è importante tenere in considerazione la zona dove sono collocate le centraline (sarebbero necessari dei dati georeferenziati) in relazione alla zona industriale dell’Ilva in continua attività, presente a Taranto.

### 4) Bibliografia

• Brunekreef B, S.T. Holgate (2002) Air pollution and health. Lancet 360.

• Fuzzi, S., 2009. Caratterizzazione del particolato fine atmosferico: attuale situazione e prospettive. In: Atti del quarto Convegno Nazionale Controllo ambientale degli agenti fisici: nuove prospettive e problematiche emergenti. Vercelli, 24-27 marzo 2009.

• Marconi A., 2003. Materiale particellare aerodisperso: definizioni, effetti sanitari, misura e sintesi delle indagini ambientali effettuate a Roma, Annali dell’Istituto Superiore di Sanità, 39(3): 329-342.