UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN SCIENZE STATISTICHE ECONOMICHE, FINANZIARE E AZIENDALI

TESI DI LAUREA

L'ozono e la salute umana: una valutazione basata sull'analisi dei dati funzionali.

Relatore: Ch.ma Prof.ssa Monica Chiogna

Laureando: Francesco Cavallin

545718 – SEFA

ANNO ACCADEMICO 2007-2008

Indice

OBIETTIVO DELLA TESI.

In questa tesi ci si propone di studiare la relazione tra l'ozono e la salute umana, utilizzando l'analisi dei dati funzionali proposta da Ramsay e Silverman (2002).

L'interesse dello comunità scientifica nei confronti dell'ozono come inquinante atmosferico è aumentato negli ultimi decenni. Infatti, studi come quelli di Clifford P. Weisel et al. (1995) e Michelle L. Bell et al. (2006) hanno evidenziato una relazione significativa tra la concentrazione di ozono e l'aumento della morbidità della popolazione esposta all'inquinante.

Gli indicatori di esposizione usualmente impiegati in letteratura per misurare l'eventuale associazione tra ozono e salute, come la concentrazione media giornaliera o la concentrazione massima giornaliera, non considerano le oscillazioni delle concentrazioni dell'ozono durante il giorno, che potrebbero ulteriormente chiarire alcuni aspetti dell'impatto dell'ozono sulla salute.

In questa tesi si cerca di affrontare questo problema, utilizzando l'analisi dei dati funzionali per cercare di sfruttare tutta l'informazione fornita dalle rilevazioni orarie dell'ozono. In tal modo si spera di superare alcuni limiti legati all'utilizzo di sintesi giornaliere.

I dati utilizzati si riferiscono alla città di Milano nel periodo 1998 - 2003. Per rappresentare la morbidità della popolazione esposta all'inquinante, è stato considerato il numero giornaliero di ricoveri registrati presso gli ospedali milanesi, escludendo i ricoveri programmati.

1. INQUINAMENTO E SALUTE.

1.1 L'OZONO COME INQUINANTE ATMOSFERICO

Negli ultimi due decenni è aumentato l'interesse della comunità scientifica nei confronti di un inquinante dell'aria delle zone urbane: l'ozono. Numerosi studi hanno constatato la relazione tra alti livelli di concentrazione di ozono nell'aria ed un peggioramento delle condizioni di salute della popolazione esposta; un esempio è lo studio svolto da Clifford P. Weisel et al. (1995), che ha evidenziato una relazione significativa tra la concentrazione di ozono rilevata ed il numero di ricoveri per asma registrati dal 1986 al 1990 in nove ospedali del New Jersey. Questi studi hanno portato alla definizione di valori limite di esposizione a questo inquinante; in Tabella 1 sono presentati i valori limite in vigore in Italia per l'ambiente esterno. In seguito, altri studi come quello di Michelle L. Bell et al. (2006) hanno mostrato un'effetto dannoso dell'ozono sulla salute anche a bassi livelli di concentrazione; inoltre, la quantità di inquinante a cui si può essere esposti dipende dall'andamento della concentrazione, che durante il giorno non è uniforme (questo aspetto verrà approfondito in seguito).

Le misure di esposizione presenti in letteratura, come la concentrazione media giornaliera o la concentrazione massima giornaliera, sono indicatori di sintesi e non considerano l'andamento orario dell'ozono. Al fine di rappresentare in modo più completo l'effetto dell'ozono, sono state suggerite misure di esposizione alternative.

M. Chiogna e P. Bellini (2002), dopo aver individuato una soglia per distinguere tra alta e bassa concentrazione di inquinante, hanno proposto tre misure di esposizione: l'intensità (calcolata come la differenza tra la concentrazione massima rilevata durante il giorno e il cutoff), la durata (pari al numero di rilevazioni giornaliere che superano la soglia) e l'esposizione notturna (calcolata come la media delle rilevazioni notturne). Queste misure esprimono l'esposizione all'inquinante in modo più appropriato degli indicatori di sintesi tradizionali, in quanto tengono conto di alcune caratteristiche dell'andamento giornaliero.

In questa tesi ci si propone di utilizzare tutta l'informazione fornita dalle 24 rilevazioni orarie della concentrazione dell'ozono. Per fare questo viene utilizzata l'analisi dei dati funzionali, proposta da Ramsay e Silverman (2002).

Questa tecnica ha permesso, nel nostro caso, di costruire delle funzioni giornaliere dell'ozono basate sulle ventiquattro rilevazioni orarie; le analisi successive sono state quindi effettuate sulle funzioni costruite e non più sui dati discreti, permettendo così di considerare l'intero andamento della variabile durante tutto il giorno.

AMBIENTE ESTERNO
• 110 μg/m_3 concentrazione media su 8 ore per la protezione della salute umana • 180 μg/m_3 concentrazione oraria di “attenzione” • 200 μg/m_3 concentrazione oraria da non raggiungere più di una volta al mese • 360 μg/m_3 concentrazione oraria di “allarme”

Tabella 1: Valori limite di esposizione in vigore in Italia.

1.2 L'OZONO E I SUOI EFFETTI SULLA SALUTE UMANA

1.2.1 L'OZONO

L'ozono è una delle due forme in cui l'ossigeno è presente in natura. E' un gas formato da tre atomi di ossigeno (O_3), dotato di un elevato potere ossidante. Ad elevate concentrazioni si presenta di colore azzurro, con un caratteristico odore pungente (dal greco *ozein* = odorare) ed è tossico. Questa tossicità è dovuta al fatto che la molecola è molto reattiva a causa dell'instabilità termodinamica (l'energia necessaria per rompere il legame è bassa, 163 kJ/mol).

In natura si trova in concentrazioni rilevanti ad alta quota (nella *troposfera*, tra i 30 e i 50 km di altezza), dove si decompone per assorbimento della radiazione ultravioletta proveniente dal sole: ciò ha come diretta conseguenza quella di schermare la superficie della Terra da questa intensa radiazione pericolosa per la salute degli esseri viventi. L'assenza di questa composto nella stratosfera è chiamato generalmente “buco dell'ozono”.

Nei bassi strati dell'atmosfera è presente in piccole concentrazioni, per effetto del naturale scambio con la stratosfera. Tale concentrazione può però aumentare nelle zone urbane e suburbane; qui la presenza di altri inquinanti chimici ne favorisce la formazione, con un conseguente aumento della concentrazione. Questo aumento si origina sopratutto nei mesi estivi, in concomitanza di una temperatura elevata e di un intenso irraggiamento solare.

1.2.2 IL COMPORTAMENTO DELL'OZONO

Il fatto più importante che influenza il comportamento dell'ozono è che non ha sorgenti proprie di origine antropogenica. A differenza degli inquinanti “primari” (che sono invece direttamente riconducibili a specifiche fonti di emissione), l'ozono si forma infatti come inquinante secondario, principalmente da reazioni fotochimiche che coinvolgono gli idrocarburi (HC) e gli ossidi di azoto (Nox) emessi da sorgenti antropogeniche (in particolare il traffico veicolare).

Questo determina un comportamento dell'ozono molto diverso rispetto a quello di inquinanti primari. Il monossido di carbonio (CO), ad esempio, presenta concentrazioni in un punto che sono linearmente correlate con le emissioni di CO di una sorgente vicina. Invece, le relazioni che legano le concentrazioni di O_3 con le emissioni di HC e NOx sono non lineari e possono essere sintetizzate come segue:

- variazioni di HC e Nox producono raramente uguali variazioni percentuali delle concentrazioni di O_3 ;
- le concentrazioni di O_3 sono alte nelle zone lontane dalle sorgenti rispetto a quelle più vicine;
- in certe condizioni la diminuzione delle emissioni di HC e Nox può determinare un aumento delle concentrazioni di O_3 .

Inoltre le variazioni spaziali sono molto graduali, per cui se in punto la concentrazione di O_3 risulta elevata allora è molto probabile che valori simili si verifichino in una vasta area circostante (da decine a centinaia di chilometri quadrati).

Notevole influenza hanno alcune variabili meteorologiche come la direzione e la velocità del vento, la stabilità atmosferica, la temperatura e l'intensità della radiazione solare. Il movimento delle masse d'aria provoca il rimescolamento delle sostanze emesse dalle sorgenti di HC e Nox e quindi raramente si riesce ad imputare le concentrazioni rilevate da una stazione a singole e ben determinate sorgenti. L'aumento della velocità del vento e dell'altezza di rimescolamento aumentano sia il volume d'aria in cui possono disperdersi HC e NOx sia la velocità di dispersione; ciò ha l'effetto di diminuire le concentrazioni di O_3 .

La luce del sole innesca le reazioni fotochimiche che provocano la dissociazione di NO_2 in NO, da cui ha inizio la catena di reazioni con gli idrocarburi (emessi dal traffico veicolare) che porta alla formazione dell'ozono. La temperatura influisce sulla velocità delle reazioni chimiche in oggetto; anche se la relazione è non lineare, l'effetto complessivo è un aumento dell'ozono in corrispondenza di un aumento della temperatura. A causa di queste influenze meteorologiche, le concentrazioni dell'O_3 presentano delle variazioni stagionali: i valori più elevati si verificano in estate (luglio e agosto) mentre i più bassi in inverno. Analogamente, sono presenti anche delle variazioni giornaliere con i valori massimi raggiunti nelle ore più calde della giornata (dalle 12 alle 18).

1.2.3 GLI EFFETTI DELL'OZONO SULLA SALUTE UMANA

I principali effetti dell'O_3 sono a carico delle vie respiratorie, dove si ha l'induzione di una risposta infiammatoria e alterazioni della permeabilità del rivestimento delle vie respiratorie. Queste alterazioni provocano una riduzione della funzione polmonare e la comparsa di iper-reattività

bronchiale. Tuttavia, alcuni studiosi hanno opinato la teoria secondo cui l'ozono comporterebbe un danno diretto alle mucose respiratorie, non ritenendo sufficienti le concentrazioni che si raggiungono dopo l'inalazione. Il danno riconosciuto in questi studi è di tipo indiretto: l'effetto ossidante dell'ozono (anche a basse concentrazioni) potrebbe modificare i componenti dello strato di muco che riveste le vie respiratorie, determinando l'alterazione della sua viscosità e la formazione di composti tossici secondari pro-infiammatori.
Gli effetti dell'ozono dipendono dalla concentrazione presente, dalla durata e dalla modalità di esposizione, nonché dalle differenze individuali. I soggetti sensibili sono anziani, bambini, donne in gravidanza, chi lavora o svolge attività fisica all'aperto; quelli a rischio sono persone asmatiche, con patologie polmonari o cardiache.

1.3 IL PM10 E I SUOI EFFETTI SULLA SALUTE UMANA

In questo paragrafo viene presentato un approfondimeno su un altro inquinante atmosferico, il PM10, che in seguito verrà utilizzato come variabile confondente.

1.3.1 LE POLVERI ATMOSFERICHE (PM10)

Il termine “polveri atmosferiche” (o anche “PM” dall'inglese “particulate matter”, che significa materiale particellare) viene utilizzato per indicare l'insieme eterogeneo di particelle solide e liquide che rimangono sospese nell'aria, a cause delle loro ridotte dimensioni.
È un insieme eterogeneo perchè queste particelle si differenziano tra loro per dimensione, forma, composizione chimica e processo di formazione. Sono costituite da una miscela di elementi: carbonio, fibre, silice, metalli, nitrati, solfati, composti organici e materiale inerte (spore,pollini..); questa composizione dipende dal processo di formazione delle particelle stesse e dalle ulteriore sostanze con cui sono entrate in contatto nella loro permanenza in atmosfera.
A seconda delle dimensioni sono suddivise in particelle grossolane (diametro>10 μm), particelle fini “PM10” (diametro compreso tra 10 μm e 2,5 μm) e particelle finissime (diametro<2,5 μm).
Le polveri atmosferiche possono essere di origine sia naturale che antropica; si può affermare che le polveri grossolane abbiamo prevalentemente origine naturale, mentre quelle fini origine antropica. Le prime sono riconducibili principalmente a materiale inorganico prodotto da agenti naturali (quali vento e pioggia), incendi boschivi ed emissioni vulcaniche. Le seconde hanno come sorgenti i trasporti, i processi industriali, la combustione incontrollata di residui agricoli e in generale quella relativa agli impianti di riscaldamento; possono formarsi anche in seguito a reazioni fisico-chimiche

in atmosfera con composti chimici come ossido di azoto, ossido di zolfo, ammoniaca e ozono. .In ambito urbano, quindi, le fonti di emissione sono principalmente il traffico veicolare e gli impianti di ricaldamento civili, senza dimenticare un 30-40 % di polveri fini formatesi a partire da inquinanti precursori (come già detto in precedenza).

Il tempo di permanenza in atmosfera dipende dalle dimensioni e va dalle poche ore delle particelle grossolane a qualche giorno per le particelle fini. Nella stagione invernale e in caso di particolari condizioni meteorologiche (alta pressione, stabilità atmosferica, assenza di precipitazione e prolungata inversione termica), la situazione di inquinamento da polveri può permanere per molti giorni.

L'effettiva rimozione delle polveri avviene per deposizione secca o umida (legata alle precipitazioni) e per effetto del vento, che è in grado di produrre importanti azioni di pulizia.

1.3.2 GLI EFFETTI DEL PM10 SULLA SALUTE UMANA

Per quanto riguarda gli effetti sulla salute umana, essi riguardano principalmente le vie respiratorie e dipendono dalle dimensioni delle particelle: le polveri fini penetrano nel tratto superiore delle vie aeree mentre le particelle finissime raggiungono in profondità polmoni e bronchi.

La dannosità è dovuta alla tossicità sia delle polveri sia delle sostanze assorbite dalle polveri stesse durante la permanenza in atmosfera. Infatti le particelle fini agiscono da veicolo per sostanze molto tossiche come alcuni metalli (piombo, cadmio e nichel) e alcuni idrocarburi.

Gli effetti riscontrati sono a breve termine (irritazione dei polmoni, broncocostrizione, tosse, bronchite cronica) e a lungo termine (un'esposizione a basse concentrazioni per un lungo periodo può produrre l'insorgenza di adenocarcinoma bronchiale).

1.4 MODELLI PER LO STUDIO DEGLI EFFETTI DELL'INQUINAMENTO SULLA SALUTE.

1.4.1 STUDI DI SERIE TEMPORALI

Una buona parte degli studi degli effetti a breve termine dell'esposizione all'inquinamento atmosferico è costutita da studi di serie storiche temporali. Questi studi mettono in relazione serie storiche di misure di esposizione della popolazione a determinati agenti inquinanti con serie storiche della mortalità o della morbosità relative agli individui esposti. L'obiettivo è cercare un'associazione tra le variazioni dei livelli degli inquinanti di interesse e le variazioni dell'indicatore della salute della popolazione esposta.

Per fare ciò, è importante considerare il problema del confondimento e l'effetto delle variabili ritardate.

Il primo aspetto si riferisce alla necessità di verificare la presenza di variabili confondenti e di considerarne l'effetto sulle stime. Infatti, la maggior parte delle variabili riferite alla salute e all'inquinamento presentano una variazione naturale nel corso del tempo. Inoltre, alcune variabili di tipo ambientale possono avere effetto sulle variazioni dell'indicatore della salute. Per individuare l'effettiva associazione tra l'indicatore della salute e la variabile riferita all'agente inquinante, è necessario rimuovere l'effetto confondente dovuto alle variazioni naturali e ad alcune variabili ambientali. A tal fine, alcune di queste variabili vengono introdotte nel modello come variabili confondenti.

Il secondo aspetto riguarda l'effetto ritardato di alcune variabili esplicative. Infatti alcune variabili meteorologiche (come la temperatura) ed alcuni agenti inquinanti possono avere effetto sulle variazioni dell'indicatore della salute anche a distanza di qualche giorno. Per questo motivo, vengono inseriti nel modello alcuni ritardi delle variabili esplicative.

1.4.2 MODELLI PRESENTI IN LETTERATURA

Negli studi di serie storiche temporali sugli effetti dell'inquinamento, la variabile di interesse è rappresentata dal numero di decessi o di ricoveri ospedalieri che avvengono in un intervallo di tempo. Questa è una variabile di conteggio ed il processo generatore dei dati è un processo di Poisson.

Detto λ_t il numero atteso di decessi nel giorno i, la probabilità che i decessi siano y_i è data da:

$$P_{Y_i}(y_i;\lambda_i) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \quad .$$

Il modello di regressione che può essere utilizzato per spiegare le variazioni dei decessi in funzione delle variabili esplicative è un modello lineare generalizzato di Poisson, in cui si assume:

$$\log(E(Y_i)) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_p X_{pi} \quad i = 1,\ldots,N$$

dove

- Y_i è il conteggio di decessi o ricoveri giornalieri,
- $E(Y_i)$ è il valore atteso di Y_i nel giorno i,
- X_{1i} ..,X_{pi} sono le variabili esplicative,
- β_1,..,β_p sono i coefficienti delle variabili esplicative.

In passato, questo modello è stato largamente utilizzato negli studi sulla relazione tra inquinamento e salute; tuttavia, nell'ultimo decennio, a questo modello è stato preferito un approccio semi-parametrico. In queste regressioni, i termini lineari rappresentano le variabili di interesse (gli

inquinanti) e le funzioni di lisciamento rappresentano le variabili considerate come “rumore” (il tempo ed alcune variabili meteorologiche).

Questo approccio ha portato ad una formulazione generale della regressione di questo tipo:

$$\log[E(Y_i)] = \beta_0 + \sum_{j=1}^{J} \beta_j X_{ji} + \sum_{j=J}^{P} f_j(x_{ji}, df_j)$$

dove

- X_{ji} sono le variabili di interesse,
- $f_j(\cdot, df_j)$ sono le funzioni di lisciamento e df_j rappresentano i parametri di lisciamento.

Questo modello permette di controllare con una maggiore flessibilità i fattori confondenti non lineari, come il trend, la stagionalità ed alcune variabili meteorologiche.

1.4.3 MODELLO UTILIZZATO IN QUESTA TESI

In questa tesi si è scelto di utilizzare un modello di regressione funzionale.

Questo consiste in un modello lineare in cui una o più variabili vengono considerate come dati funzionali; nel nostro caso, la variabile inclusa sotto forma funzionale è quella che rappresenta la concentrazione dell'ozono.

L'indicatore della salute della popolazione è rappresentato dal numero dei ricoveri giornalieri; poiché questa è una variabile di conteggio, al fine di riportare i dati ad una condizione di normalità, nel modello sono state considerate due sue trasformazioni, ovvero:

$$y_{iA} = \log(r_i)$$

e

$$y_{iB} = \overline{r_i} - \left[\overline{r_i - \frac{1}{2}}\right]$$

dove r_i rappresenta il numero dei ricoveri giornalieri nel giorno i.

La variabile esplicativa principale è l'ozono, di cui si vuole studiare la relazione con la morbidità della popolazione esposta. Questa variabile è stata inserita nel modello come dato funzionale, al fine di considerarne anche l'andamento non uniforme durante il giorno.

Oltre all'ozono, sono state considerate come variabili esplicative anche alcune variabili confondenti. Tra queste vi sono alcune variabili temporali come l'indicatore del giorno festivo e l'indicatore del giorno della settimana. L'indicatore di giorno festivo ha lo scopo di cogliere una variazione del traffico cittadino (responsabile dell'aumento della concentrazione dell'ozono) ed una eventuale maggiore espozione della popolazione all'inquinante, dovuta al maggior tempo passato all'aperto. L'indicatore del giorno della settimana esprime le possibili variazioni dovute alle attività della popolazione, che si svolgono in maniera differente a seconda del giorno. Inoltre, poiché molte variabili riferite alla salute ed all'inquinamento presentano una variazione naturale nel corso del

tempo, nel modello è stato considerato anche un indicatore temporale progressivo con lo scopo di cogliere questo trend naturale.

Tra le variabili confondenti sono presenti anche due variabili climatiche e un'altro agente inquinante. Quest'ultimo è il pm10, che è un inquinante atmosferico che causa disturbi respiratori come l'ozono e pertanto è necessario considerarlo come variabile confondente. Nel modello è considerata la sua concentrazione media giornaliera.

Una delle variabili climatiche è la temperatura media giornaliera; infatti, numerosi studi hanno evidenziato una relazione tra variazioni della temperatura e variazioni dello stato di salute della popolazione. Inoltre, è stata considerato rilevante anche l'effetto ritardato della temperatura; per questo motivo è stata inserita nel modello anche una variabile che esprime la differenza tra la temperatura corrente e quella media dei tre giorni precedenti.

La formulazione generale di questo modello è la seguente:

$$Y_i = \beta_0 + \beta_1(t)x_{i1}(t) + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\dots,711 \quad t = 1,..,24$$

dove:

- i indica il giorno e t l'ora,
- Y_i rappresenta una trasformazione del numero dei ricoveri,
- $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale,
- $x_{i2},\dots,x_{i7}$ sono le variabili confondenti,
- $\beta_1(t)$ è il coefficiente di regressione dell'ozono, espresso in forma funzionale,
- $\beta_2,\dots,\beta_7$ sono i coefficienti di regressione delle variabili confondenti,
- ε_i è il termine di errore.

Si assume che i termini di errore siano indipendenti, abbiano media nulla ed abbiano varianza comune.

L'ozono in forma funzionale è espresso dalla formula

$$x_{i1}(t) = \sum_{k=1}^{K} c_{ik}\phi_k$$

dove c_{ik} sono i coefficienti reali e ϕ_k le funzioni di base.

In modo analogo, il coefficiente di regressione dell'ozono è espresso in forma funzionale da

$$\beta_1(t) = \sum_{k=1}^{K} \beta_k \phi_k$$

dove β_k sono i coefficienti reali e ϕ_k le funzioni di base.

Il concetto di forma funzionale e di funzioni di base verrà esposto nel capitolo successivo.

2. L'ANALISI DEI DATI FUNZIONALI.

2.1 COSA SONO I DATI FUNZIONALI

In alcune analisi statistiche, per ogni unità statistica i dati disponibili sono costituiti da più osservazioni nel tempo della stessa variabile, come avviene, per esempio, nel caso delle misure ripetute. Ogni unità statistica è rappresentata in questo caso da una o più serie temporali di valori riferiti ad una o più variabili. Se le rilevazioni temporali sono sufficientemente frequenti, si può pensare di identificare le funzioni temporali sottostanti l'andamento dei dati e di utilizzare tali funzioni nelle analisi successive. In questo modo, l'informazione relativa ad ogni unità statistica non è più una serie discreta di valori, ma è una funzione continua. Ciò permette di analizzare alcune proprietà, come la velocità (rappresentata dalla derivata prima) e l'accelerazione (rappresentata dalla derivata seconda).

Le funzioni ricavate dai dati vengono chiamate "dati funzionali".

I metodi per analizzare questo tipo di dati sono detti "metodi di analisi dei dati funzionali" (acronimo: FDA).

Ramsay e Silverman presentano alcuni esempi efficaci di dati funzionali nel testo "Functional Data Analysis" (2002), due dei quali verranno qui brevemente riportati a scopo illustrativo.

Un esempio riguarda un campione di 10 ragazzi svizzeri, di cui viene misurata l'altezza per analizzare l'accelerazione nella crescita in età puberale. A tal fine, ad ogni ragazzo viene misurata l'altezza a determinate età per un totale di 29 misurazioni. Si ottiene quindi per ciascuno di essi una curva della crescita in funzione dell'età. L'insieme complessivo dei dati, costituito da 290 osservazioni, può essere così visto come un campione di 10 funzioni di crescita. Questo permette di studiare le derivate seconde delle funzioni e di poter rappresentare l'accelerazione nella crescita. Il risultato identifica due fasi di accelerazioni nella crescita (intorno ai 2 e ai 6 anni) seguiti da una brusca decelerazione intorno ai 14 anni di età.

Un altro esempio riguarda le rilevazioni giornaliere della temperatura e delle precipitazioni effettuate nel corso di un anno presso 35 stazioni meteorologiche in Canada. Nel corso dell'analisi, queste stazioni vengono poi divise in 4 gruppi che identificano le diverse zone climatiche del Canada, poiché uno degli obiettivi degli autori è analizzare l'andamento della temperatura durante l'anno in funzione della zona climatica. Per ognuna delle 35 stazioni si dispone pertanto di una funzione della temperatura basata su 365 rilevazioni giornaliere e quindi l'insieme dei dati è considerato come un campione di 35 funzioni della temperatura.

Da un punto di vista pratico, i dati funzionali vengono osservati e memorizzati in modo discreto, non essendo possibile farlo nel continuo. Si parla comunque di funzioni, basate sui dati, che si possono teoricamente valutare ad ogni istante. Per fare ciò, è necessario disporre di una qualche forma di interpolazione o lisciamento dei valori discreti. A tal fine vengono utilizzate delle particolari strutture di funzioni denominate “basi”. Nel paragrafo seguente vengono introdotte le basi ed i loro principali tipi.

2.2 LE BASI

Una base è un sistema di funzioni che vengono combinate linearmente per approssimare una funzione.

I dati funzionali possono quindi essere rappresentati come combinazioni lineare di K funzioni di base:

$$x(t) = \sum_{k=1}^{K} c_k \phi_k \quad , \quad k = 1, \ldots, K$$

dove c_k sono i coefficienti reali e ϕ_k le funzioni di base.

La conversione dei dati in forma funzionale implica il calcolo dei coefficienti, che vengono determinati utilizzando il criterio dei minimi quadrati.

Le funzioni di base da preferire sono quelle che rispecchiano le caratteristiche appartenenti alla funzione da stimare: ad esempio, per una variabile con andamento periodico sono consigliabili le funzioni seno e coseno come funzioni di base. Una base deve essere scelta con lo scopo di raggiungere una buona approssimazione utilizzando però un basso valore di K: questo è importante sia dal punto di vista computazionale sia da quello dell'interpretazione. Non esiste una base universale, è necessario individuarla a partire dai dati.

In questa tesi sono stati utilizzati tre tipi di base, ritenuti a priori i più adatti a descrivere le caratteristiche della variabile funzionale considerata.

2.2.1 LA BASE COSTANTE

La base costante è costituita da una singola funzione di base che assume il valore 1 ovunque, per cui in pratica non determina alcuna trasformazione; viene usata per definire le funzioni costanti e per convertire le osservazioni scalari univariate in forma funzionale.

2.2.2 LA BASE FOURIER

La base di Fourier è una delle più note ed utilizzate; è definita a partire dalla serie di Fourier:

$$x(t) = c_0 + c_1 \phi_1(t) + c_2 \phi_2(t) + c_3 \phi_3(t) + \ldots$$

con le seguenti componenti:

$$\phi_0(t) = 1 \quad ,$$

$$\phi_{2r-1}(t) = sen(r\omega t) \quad ,$$

$$\phi_{2r}(t) = \cos(r\omega t) \quad .$$

La base è periodica ed il termine ω determina il periodo 2π/ω, che rappresenta l'intervallo di tempo su cui si lavora. Pertanto, questo tipo di base è indicata per dati con andamento periodico e che non presentano eccessive variazioni locali, mentre non è appropriata se si sospetta che la funzione dei dati presenti delle discontinuità.

2.2.3 LA BASE BSPLINE

Come è stato detto, la base di Fourier non riesce a cogliere le variazioni locali: per questo sono stati sviluppati le "splines" polinomiali. Queste sono delle funzioni costruite utilizzando dei polinomi uniti in corrispondenza di determinati valori τ_k , detti nodi.

Il numero di nodi, indicato con K_1+1, ed i nodi esterni definiscono l'intervallo su cui si lavora; tra due nodi adiacenti una "spline" polinomiale è un polinomio di grado K_2 ; in corrispondenza di un nodo interno, è richiesto che i polinomi adiacenti assumano gli stessi valori per un fissato numero di derivate (di solito K_2-1). Di solito viene scelto un grado dei polinomi pari a tre, in modo da garantire la continuità delle prime due derivate e ottenere una curva liscia.

Un modo semplice di rappresentazione è una combinazione lineare delle funzioni di base

$$\phi_k(t) = (t - \tau_k)^{K_2} \quad ,$$

che assumono valore nullo se la quantità tra parentesi è negativa.

La base bspline è un caso particolare degli "splines" polinomiali: infatti le "bsplines" sono uguali a zero ovunque tranne su un intervallo finito; nel caso degli "splines" cubici, ad esempio, ogni bspline è una spline cubica con valori sull'intervallo $[\tau_{k-2},\tau_{k+2}]$.

Il vantaggio di questo tipo di base sta nella possibilità di combinare la semplicità computazionale dei polinomi alla capacità di rappresentare le variazioni locali.

3. I DATI.

3.1 IL PERIODO TEMPORALE.

I dati utilizzati in questa tesi sono stati ricavati dalle rilevazioni effettuate nella città di Milano dal 1998 al 2003. L'attenzione è stata focalizzata sui mesi più caldi, da maggio a settembre, poichè l'ozono è un inquinante estivo.

Come si è visto nel capitolo 2, la concentrazione dell'ozono aumenta in concomitanza di una temperatura elevata e di un intenso irraggiamento solare e proprio in questi mesi la radiazione solare è più intensa.

3.2 LA VARIABILE RISPOSTA

La variabile risposta è rappresentata dal numero dei ricoveri giornalieri presso gli ospedali di Milano; sono stati esclusi i ricoveri programmati, considerati non rilevanti al fine di questa tesi.

Poiché questa è una variabile di conteggio, al fine di riportare i dati ad una condizione di normalità, nel modello sono state considerate due sue trasformazioni, ovvero:

$$y_{iA} = \log(r_i)$$

e

$$y_{iB} = \sqrt{r_i} + \sqrt{r_i + \frac{1}{2}}$$

dove r_i rappresenta il numero dei ricoveri giornalieri nel giorno i.

3.3 LE VARIABILI ESPLICATIVE.

Le variabili utilizzate come esplicative sono:

1. l'ozono (di cui vengono utilizzate le rilevazioni orarie),
2. la temperatura media giornaliera,
3. la concentrazione di pm10 (espressa come media giornaliera),
4. un indicatore dei giorni festivi,
5. una variazione di temperatura (rappresentata dalla differenza tra la temperatura media giornaliera e la media delle temperature dei tre giorni precedenti),
6. un indicatore numerico per il giorno della settimana (domenica=0, lunedì=1,...,sabato=6)
7. un indicatore progressivo giornaliero (che assume valori compresi tra 0 e 1).

La temperatura è presente nei modelli che considerano gli effetti degli inquinanti sulla salute e inoltre è correlata con l'ozono, con un aumento della concentrazione di quest'ultimo in concomitanza di temperature elevate. In questo lavoro, oltre alla misura giornaliera, ne viene considerata anche la variazione rispetto ai giorni precedenti.
Il pm10 è un inquinante atmosferico che causa disturbi respiratori come l'ozono e può quindi essere utilizzato come variabile confondente.
L'indicatore di vacanza ha lo scopo di cogliere una variazione del traffico cittadino (responsabile dell'aumento della concentrazione dell'ozono) ed una eventuale maggiore espozione della popolazione all'inquinante, dovuta al maggior tempo passato all'aperto durante i giorni di vacanza.
L'indicatore del giorno della settimana esprime le possibili variazioni dovute alle attività della popolazione, che si svolgono in maniera differente a seconda del giorno.
L'indicatore progressivo giornaliero ha lo scopo di isolare un trend nel tempo, permettendo di valutare l'effetto delle altre variabili in modo più efficace.

3.4 L'ANALISI DEI DATI.

L'analisi dei dati è stata svolta con il programma statistico R (versione 2.5.1) ed in particolare utilizzando la libreria FDA, appositamente realizzata per l'analisi dei dati funzionali.
I dati qui utilizzati si riferiscono ad un periodo di 6 anni, dal 1998 al 2003; in particolare, si è deciso di considerare solo i mesi più caldi ed è stato scelto un intervallo di tempo totale di 911 giorni. Le rilevazioni orarie dell'ozono sono contenute in un vettore di lunghezza 21864, ovvero 24 misure per 911 giorni.
Questo vettore è stato trasformato in una matrice con 24 righe e 911 colonne (che rappresentano, rispettivamente, le ore e i giorni) per poter essere trasformato in un “oggetto funzionale”. Le altre variabili, disponendo di una sola osservazione giornaliera, sono contenute in vettori di lunghezza 911.

3.4.1 I DATI MANCANTI.

I dati mancanti relativi a ozono, temperatura e pm10 hanno rappresentato un problema nell'elaborazione: quasi tutte le funzioni usate della libreria FDA del programma R restituivano un errore in presenza di valori mancanti. Si è resa pertanto necessaria la loro sostituzione; a tal scopo si è scelto di utilizzare il rispettivo valore mediano.

3.4.2 ANALISI DESCRITTIVE E GRAFICHE.

In questo paragrafo vengono presentate alcune analisi descrittive delle variabili di interesse.
Il numero dei ricoveri ha una mediana di 42 ricoveri giornalieri, con un intervallo interquartile di 36 – 48 ricoveri giornalieri; in Figura 1 è rappresentato l'andamento nel tempo di questa variabile: si possono osservare due picchi che raggiungono un centinaio di ricoveri (uno nel 2002 e l'altro nel 2003) e un valore minimo che indica 16 ricoveri (nel 2002), ma a parte queste eccezioni si nota un andamento con un'evidente caratteristica periodica ed un leggero trend crescente.

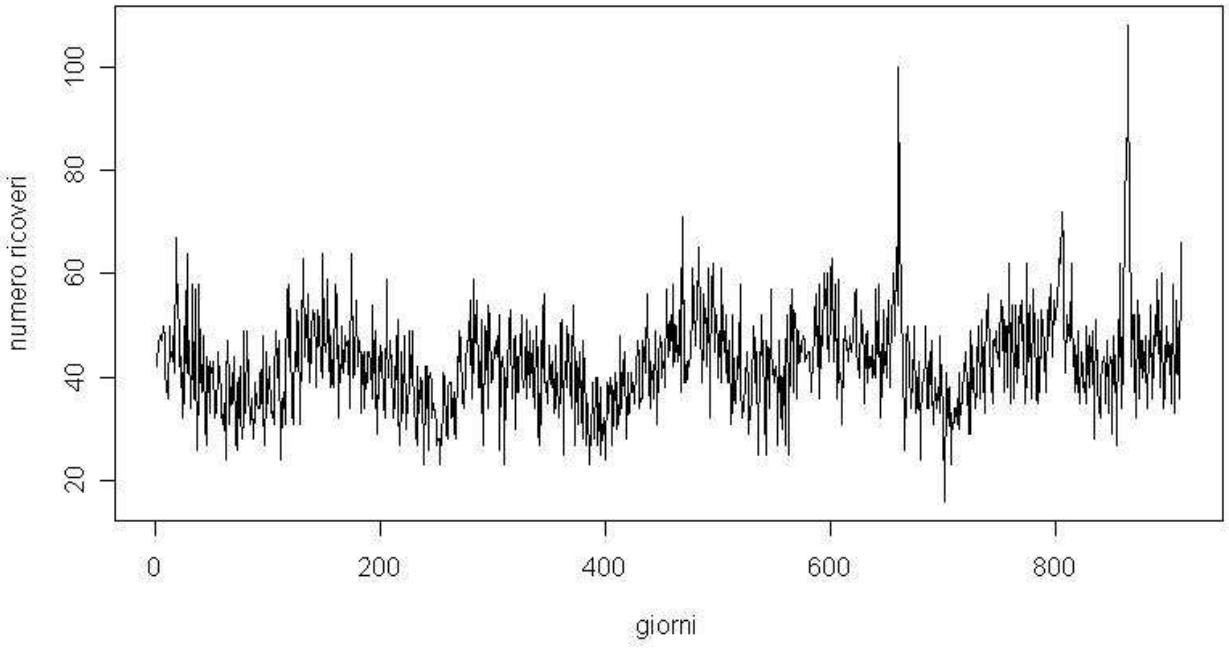

Figura 1: Andamento nel tempo del numero dei ricoveri.

Considerando la trasformazione logaritmica del numero dei ricoveri (Figura 2) e la seconda trasformazione (Figura 3) sono ancora ovviamente evidenti i due picchi ed il minimo già osservati in precedenza; anche in questi due casi l'andamento mostra una caratteristica periodica ed un leggero trend crescente.

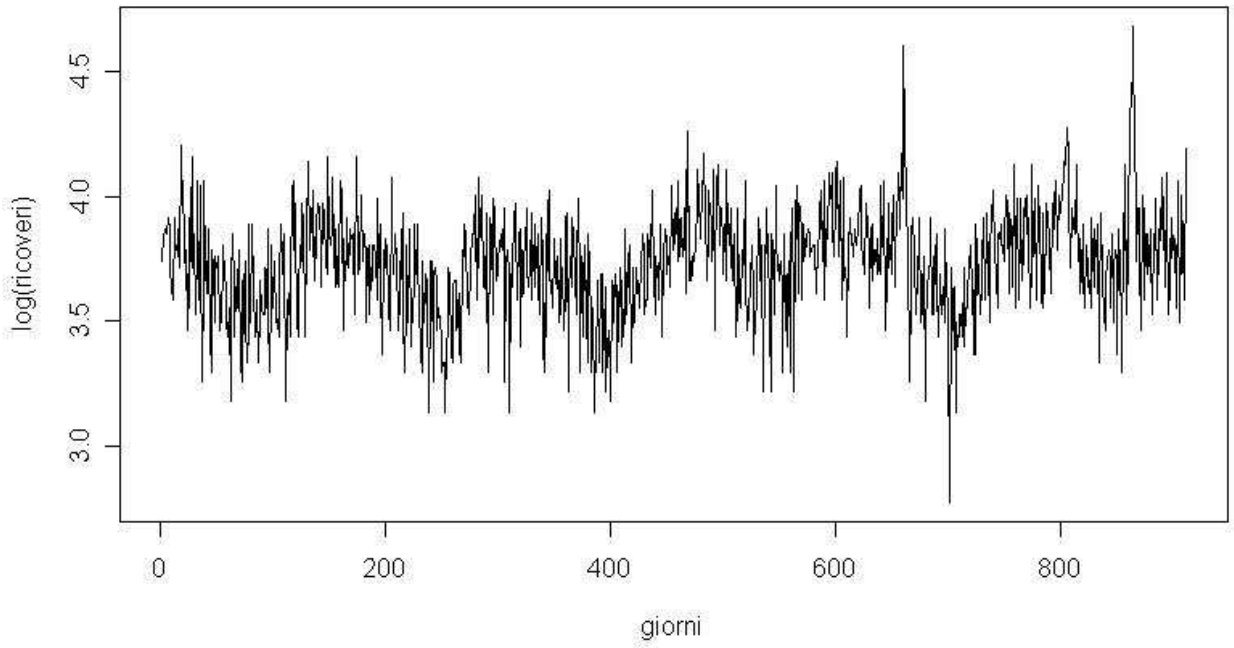

Figura 2: Andamento nel tempo di log(ricoveri).

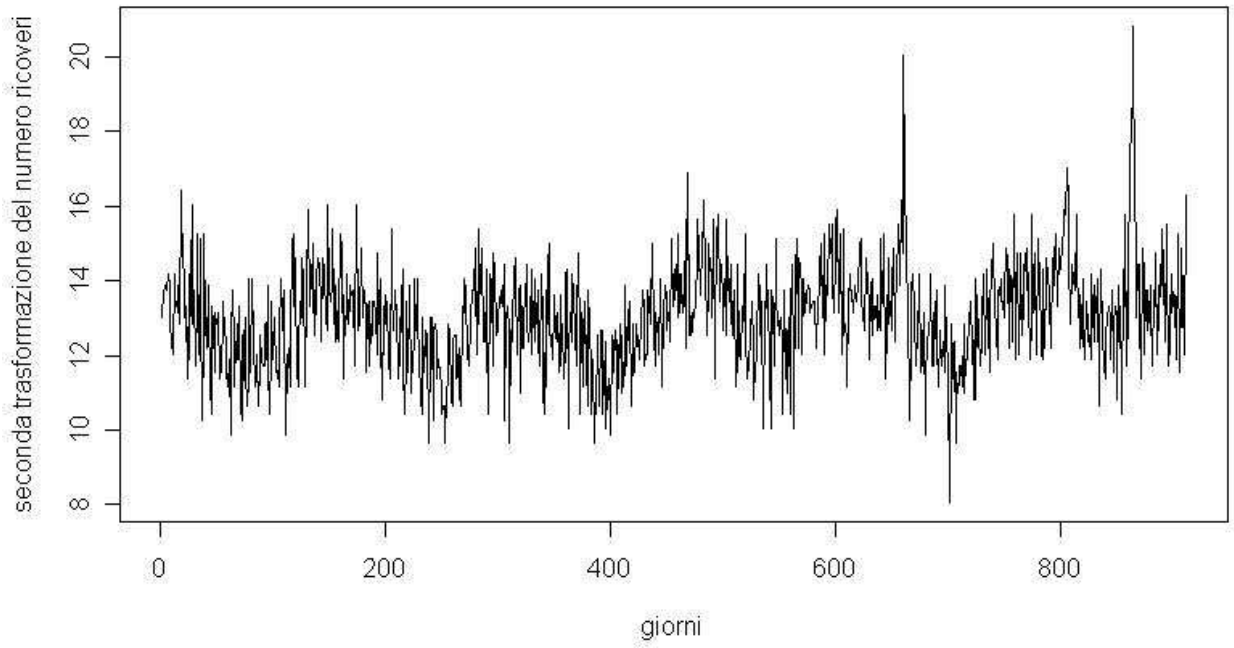

Figura 3: Andamento nel tempo della seconda trasformazione del numero di ricoveri

Anche l'ozono mostra un andamento periodico, come si può vedere dalla Figura 4, che evidenzia anche un aumento della variabilità. La concentrazione mediana rileva un livello di ozono pari a 67.77 $\mu g/m^3$ con un intervallo interquartile di 50.14 - 84.43 $\mu g/m^3$.
La Figura 5, dove sono mostrati i diagramma a scatola della concentrazione di ozono in funzione dell'anno, sembra confermare quanto detto in precedenza: i valori sono più dispersi negli ultimi tre anni, seppure la mediana e lo scarto interquartile non siano molto differenti nei vari anni.

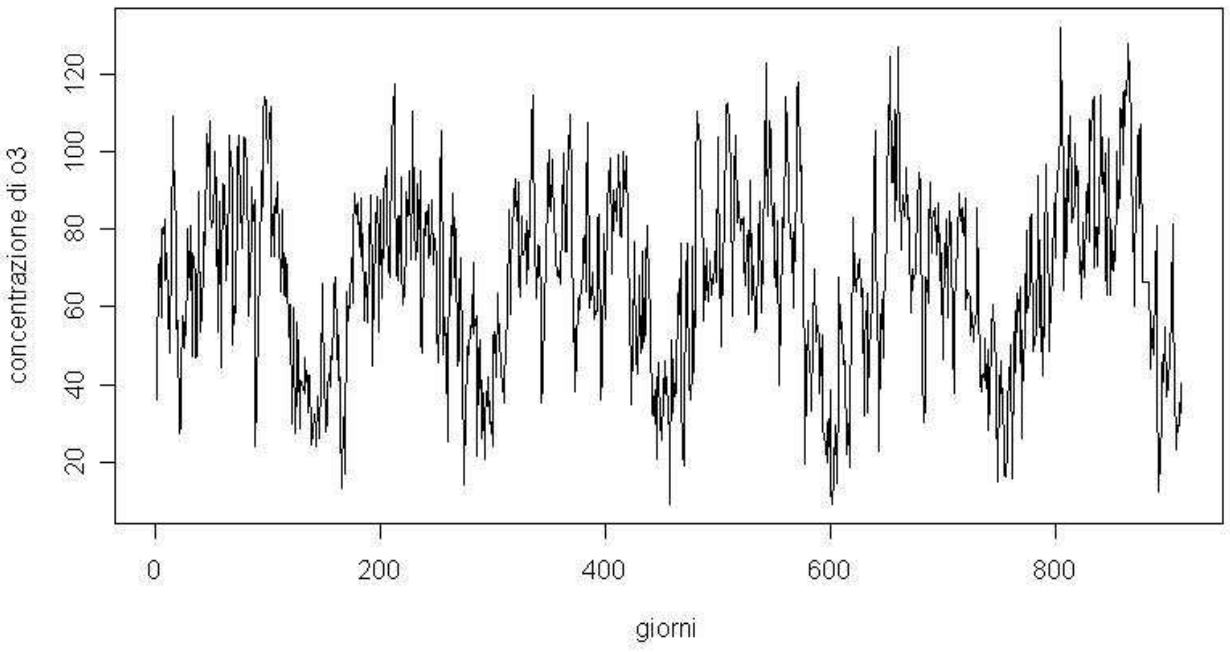

Figura 4: Andamento nel tempo dell'ozono.

Una differenza si nota invece quando si rappresenta la concentrazione dell'ozono in funzione del mese: la Figura 6 mostra chiaramente che nei mesi più caldi (giugno, luglio, agosto) le concentrazioni rilevate sono mediamente più elevate, mentre a settembre si osservano concentrazioni più basse. Questa situazione conferma quanto previsto dalla letteratura, poiché l'aumento della concentrazione di ozono è legato all'aumento della temperatura e ad un intenso irraggiamento solare, condizioni tipiche dei mesi estivi.
Durante i giorni feriali i valori della concentrazione di ozono non mostrano variazioni apprezzabili, mentre nei giorni non feriali (sabato e domenica) si attestano su livelli superiori rispetto agli altri giorni della settimana. La spiegazione di ciò potrebbe dipendere dalla formazione dei precursori dell'ozono derivanti dal traffico veicolare (come gli ossidi di azoto) durante i giorni feriali, il che determina un livello più elevato di ozono a distanza di poco tempo; per lo stesso motivo, durante il

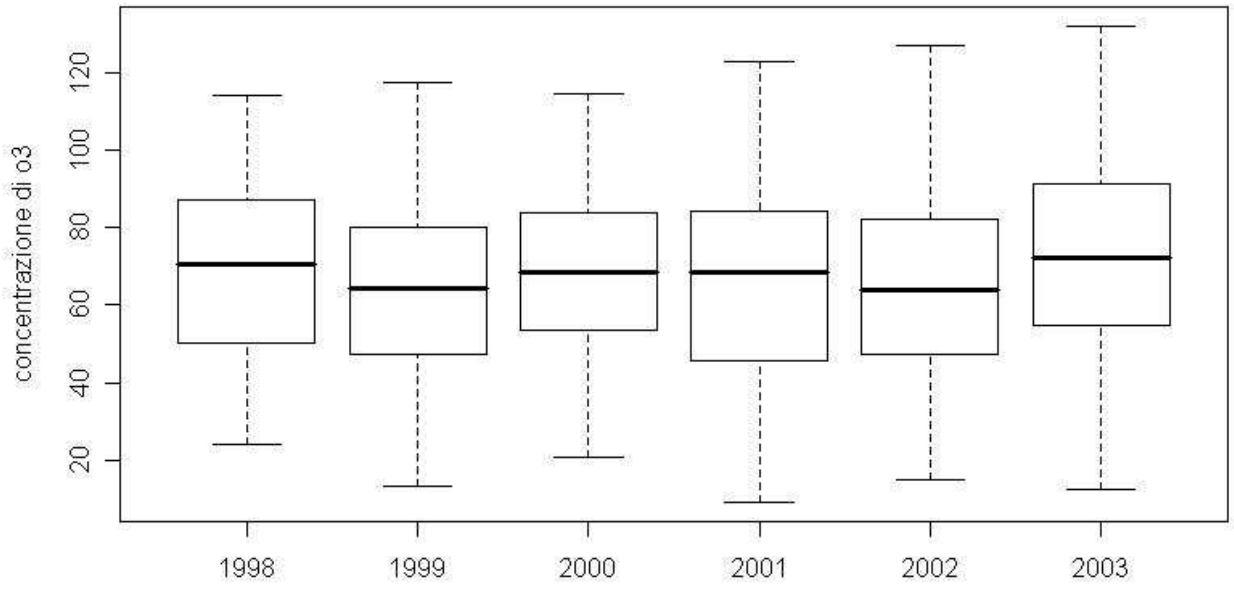

Figura 5: Diagramma a scatola dell'ozono in funzione dell'anno.

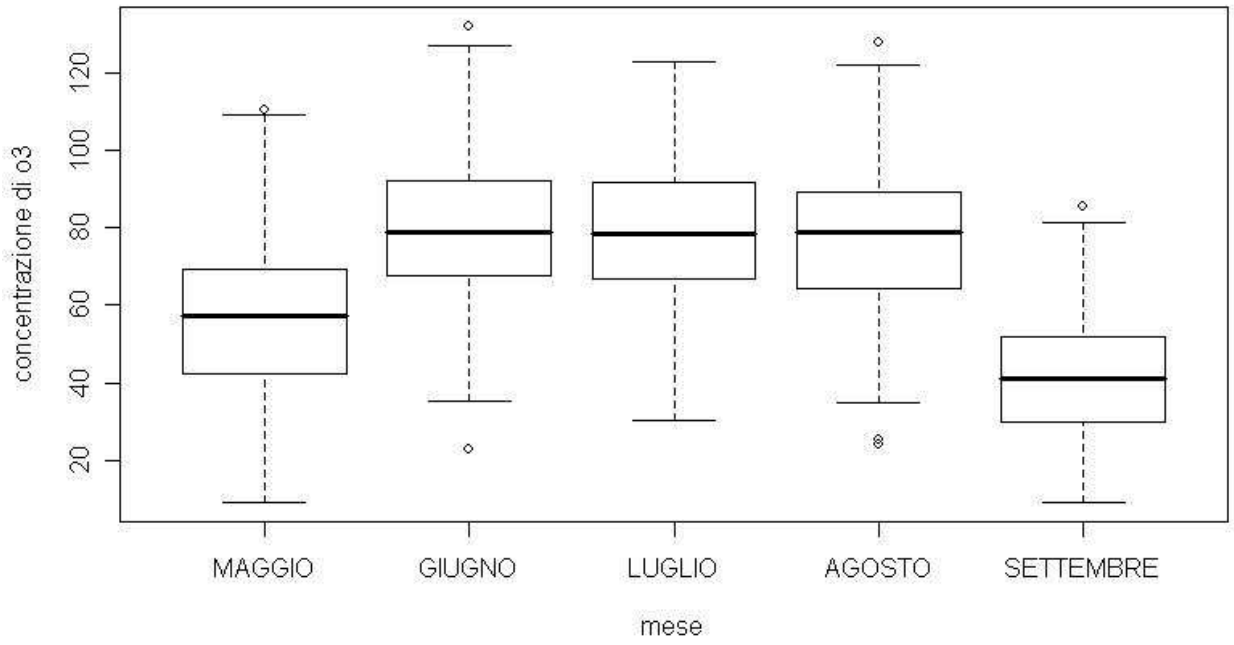

Figura 6: Diagramma a scatola dell'ozono in funzione del mese.

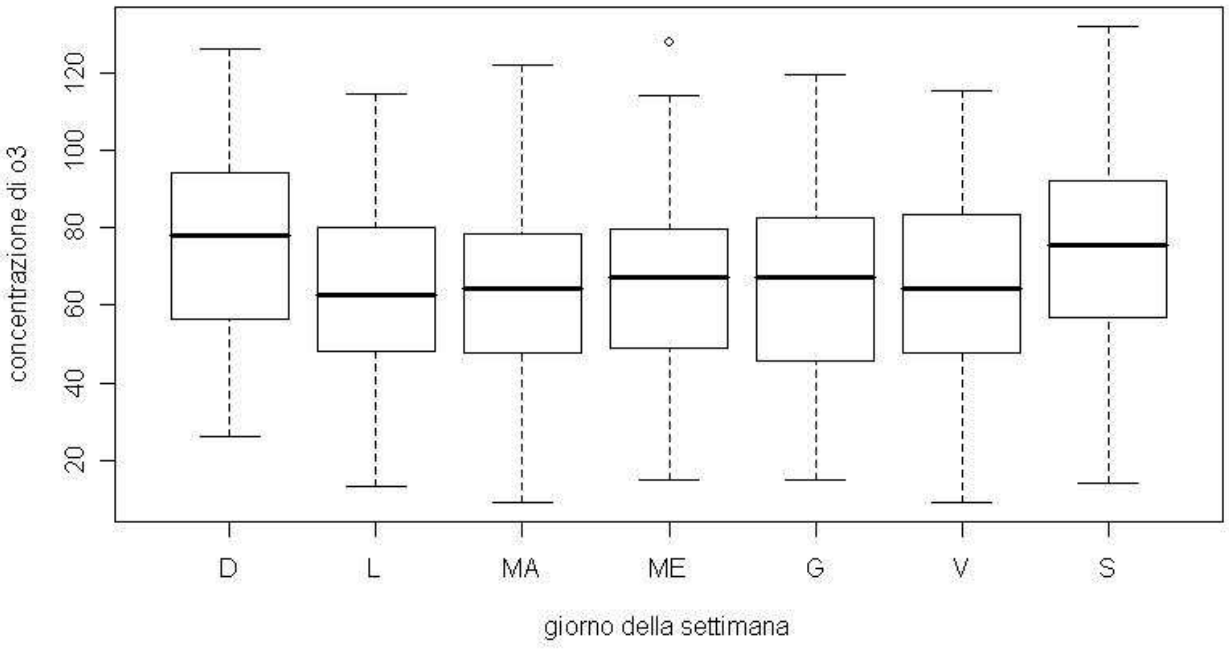

Figura 7: Diagramma a scatola dell'ozono in funzione del giorno della settimana.

fine settimana, la riduzione del traffico veicolare in città determina una diminuzione di questi precursori, il che porta ad un abbassamento della concentrazione di ozono riscontrabile a partire dal lunedì successivo. Questa ipotesi si basa sul comportamento dell'ozono descritto in letteratura, la cui variazione in funzione dei precursori risulta ritardata nel tempo e dilatata nello spazio.

In Figura 8 viene mostrato l'andamento nel tempo del PM10: non sono evidenti né una componente periodica né un trend, come si può notare anche dalla Figura 9, dove sono rappresentati i diagramma a scatola del pm10 in funzione dell'anno. A parte i valori leggermente inferiori relativi agli anni 2000 e 2001, non si osservano differenze rilevanti tra i vari anni. La mediana calcolata sull'intero periodo è pari a 33.61 $\mu g/m^3$ con un intervallo interquartile di 24.98 – 42 $\mu g/m^3$.

Anche per quanto riguarda i mesi, non si riscontrano variazioni rilevanti (Figura 10); solo il mese di settembre fa rilevare un valore mediano più elevato e una maggiore variabilità.

Se invece si considerano i valori assunti da questo inquinante atmosferico a seconda dei giorni (Figura 11), si può notare una precisa variazione: il valore del PM10 aumenta da lunedì a mercoledì, rimane ad alti livelli giovedì e venerdì, quindi diminuisce a partire da sabato per raggiungere il valore mediano minimo domenica. Questo comportamento settimanale è causato dal traffico veicolare cittadino, che è più intenso nei giorni lavorativi mentre si riduce notevolmente sabato e domenica.

Anche per l'ozono era stata osservata una variazione durante la settimana, ma la differenza risiede nella tempestività della manifestazione: mentre l'ozono è un inquinate secondario (si forma a partire

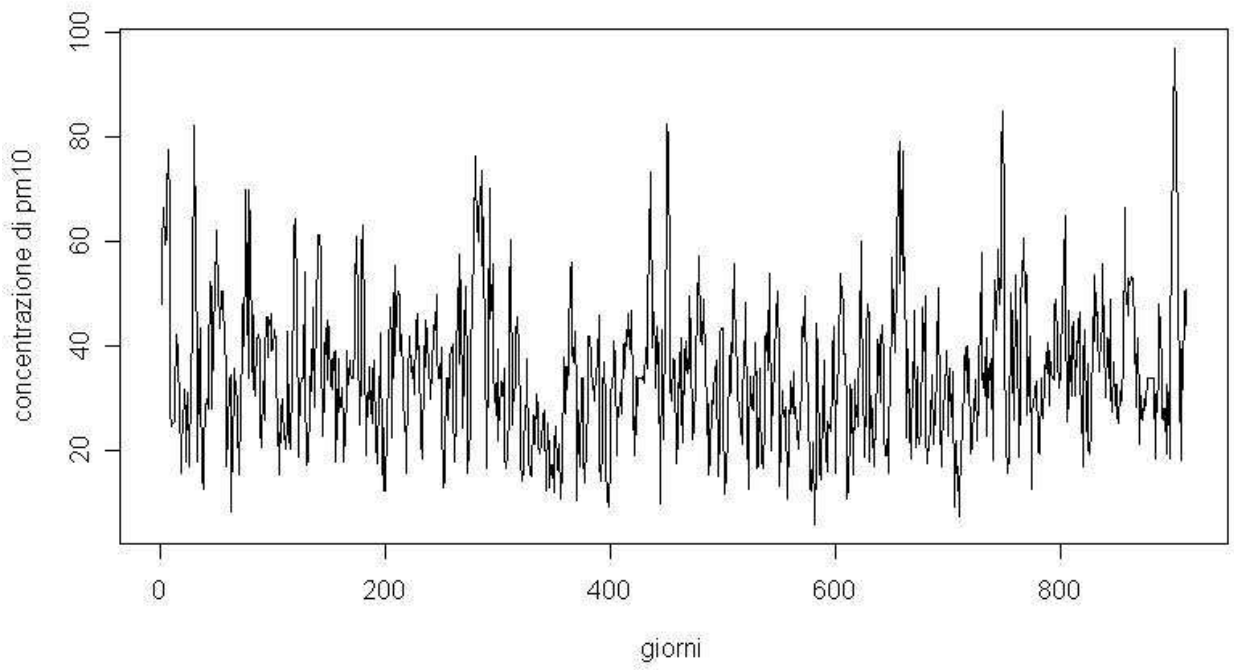

Figura 8: Andamento del PM10 nel tempo.

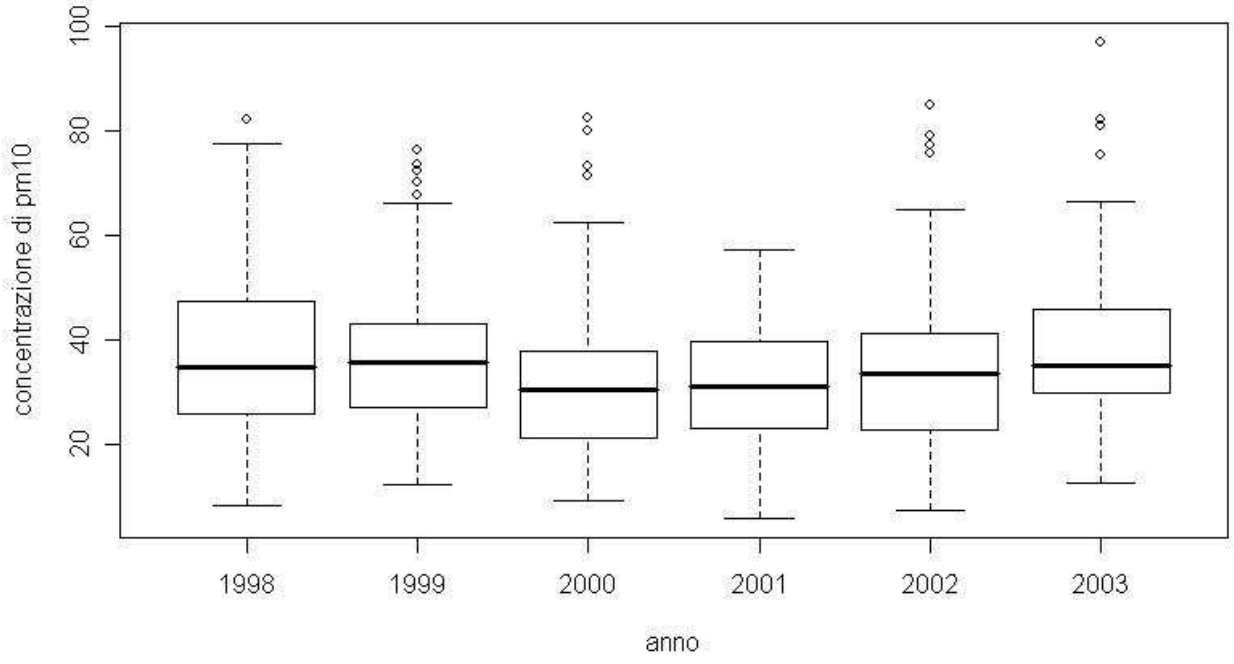

Figura 9: Diagramma a scatola del PM10 in funzione dell'anno.

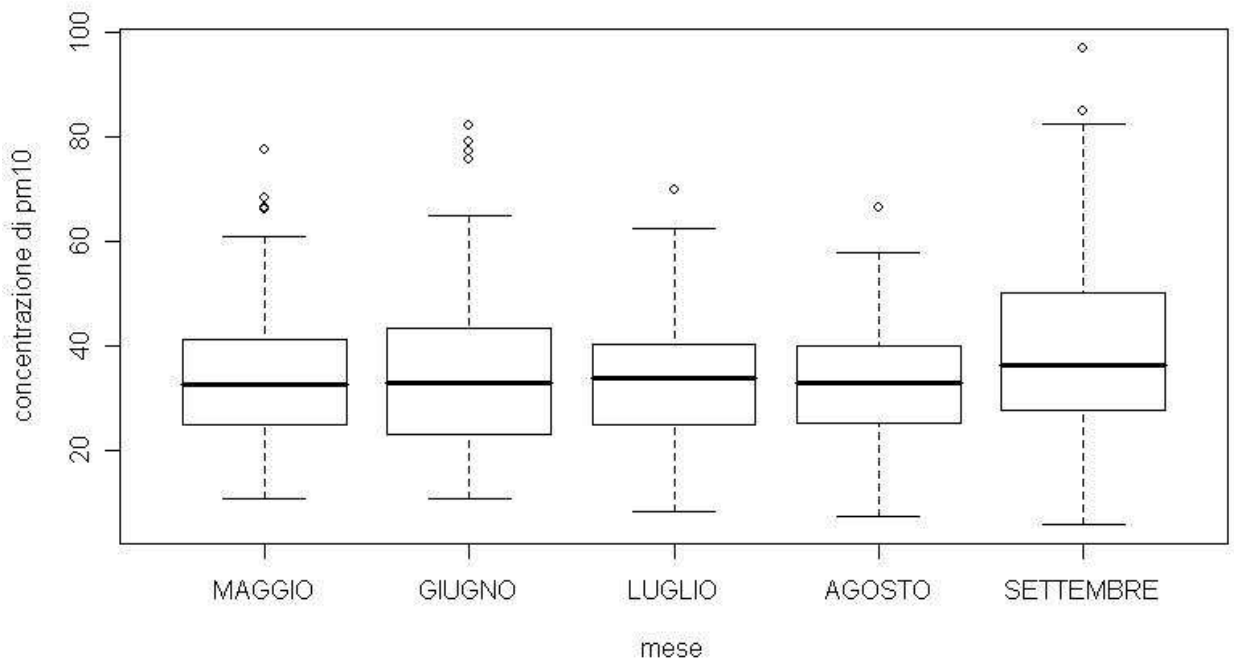

Figura 10: diagramma a scatola del PM10 in funzione del mese.

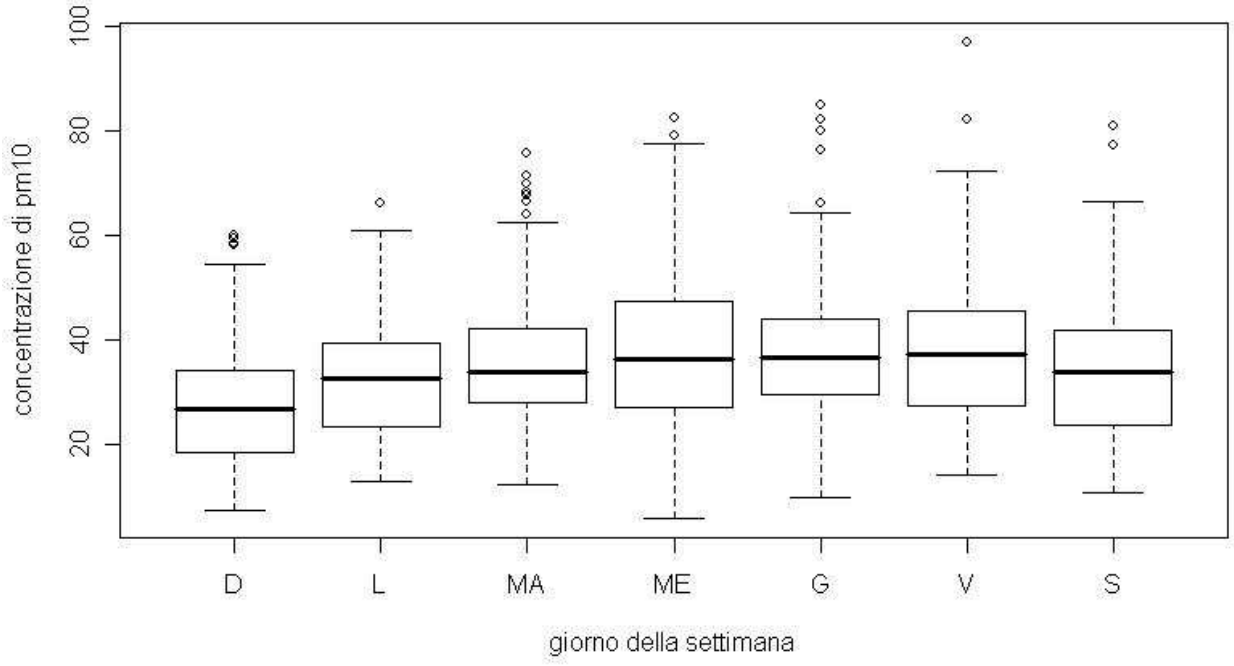

Figura 11: Diagramma a scatola del PM10 in funzione del giorno della settimana.

da inquinanti primari detti “precursori”) e quindi un suo aumento si manifesta a qualche giorno dal fenomeno precursore, il PM10 è un inquinante primario (è causato direttamente dalle emissioni dei veicoli) e pertanto la sua variazione si riscontra entro breve tempo dalla variazione del traffico.
Consideriamo ora le due variabili relative alla temperatura.
In Figura 12 viene mostrato l'andamento nel tempo della temperatura media giornaliera: i valori si

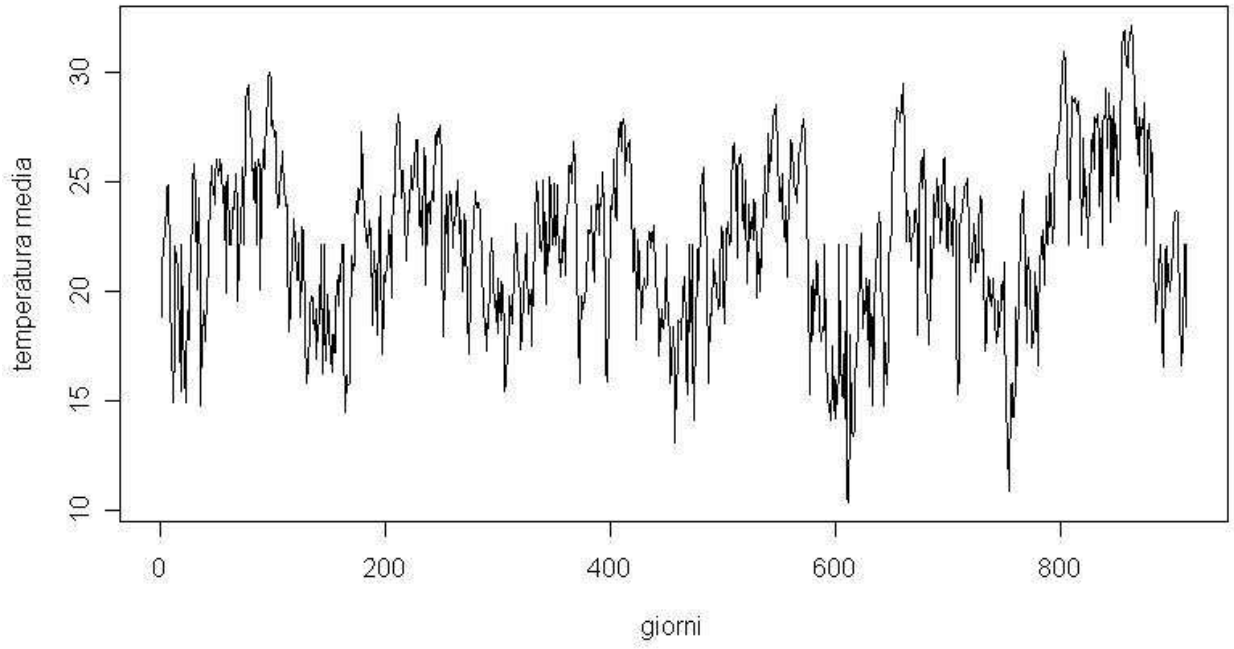

Figura 12: Andamento della temperatura media giornaliera nel tempo.

presentano con un'ovvia struttura periodica annuale e non sembra che ci sia la presenza di un trend. Nel 2003 i valori registrati sono più elevati degli altri anni, mentre nel 2001 e nel 2002 si osservano i minimi del periodo di 6 anni in esame, come si può vedere anche dalla figura 13 che riassume queste informazioni a livello annuale; a parte questo, tuttavia, si può affermare che la temperatura media annuale non abbia subito variazioni rilevanti dal 1998 al 2002. La mediana sull'intero periodo è pari a 22.25 °C con un intervallo interquartile di 19.70 – 24.86 °C.
A livello mensile, la temperatura si comporta come ci si aspetterebbe: aumenta da maggio ad agosto e fa registrare livelli più bassi nel mese di settembre (Figura 14).
Se si considerano i valori in funzione del giorno della settimana, invece, la temperatura si mantiene sugli stessi livelli e non sono osservabili variazioni rilevanti tra i vari giorni, come si può vedere in Figura 15.

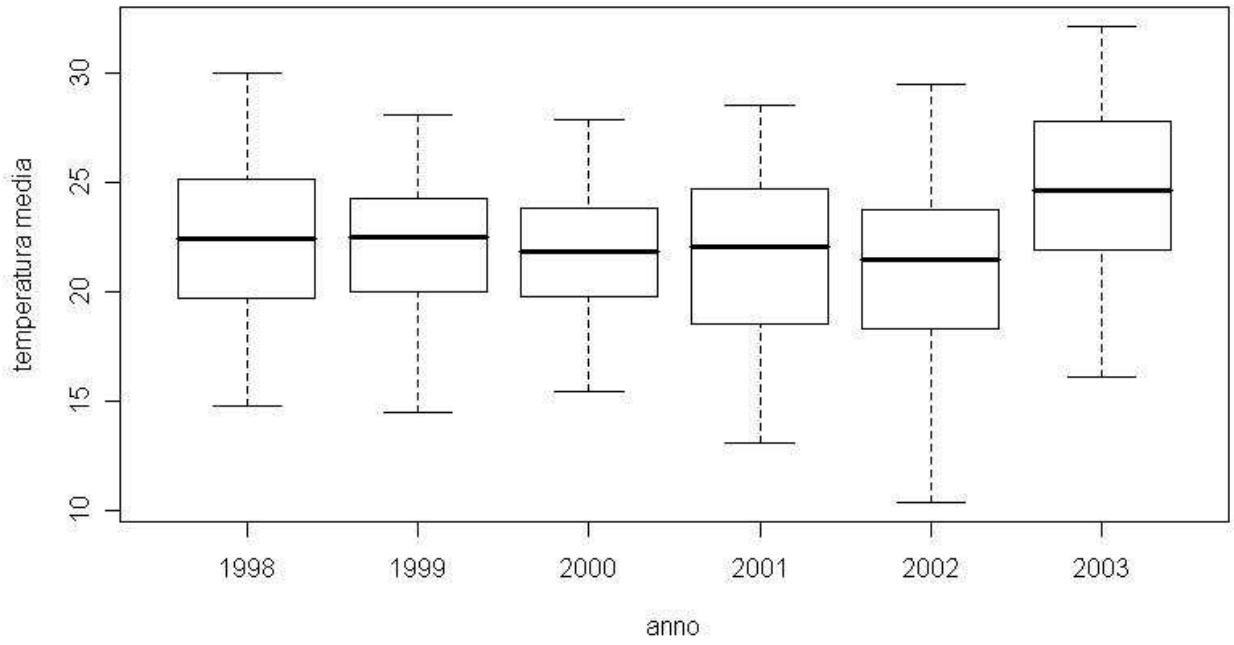

Figura 13: Diagramma a scatola della temperatura media giornaliera in funzione dell'anno.

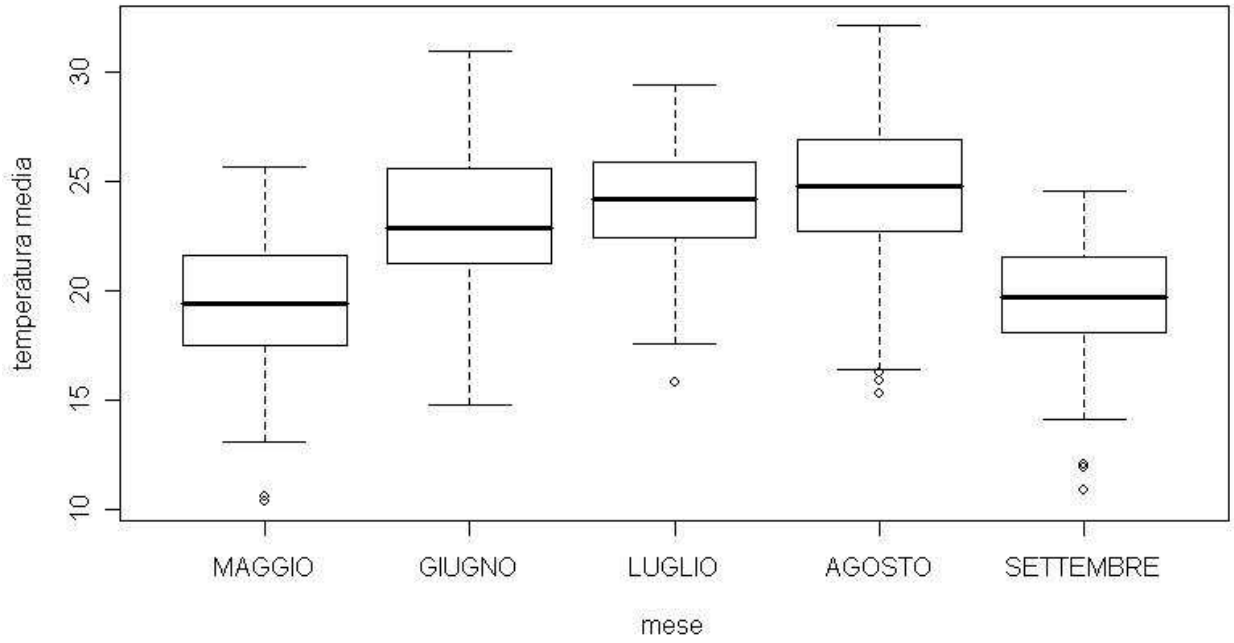

Figura 14: Diagramma a scatola della temperatura media giornaliera in funzione del mese.

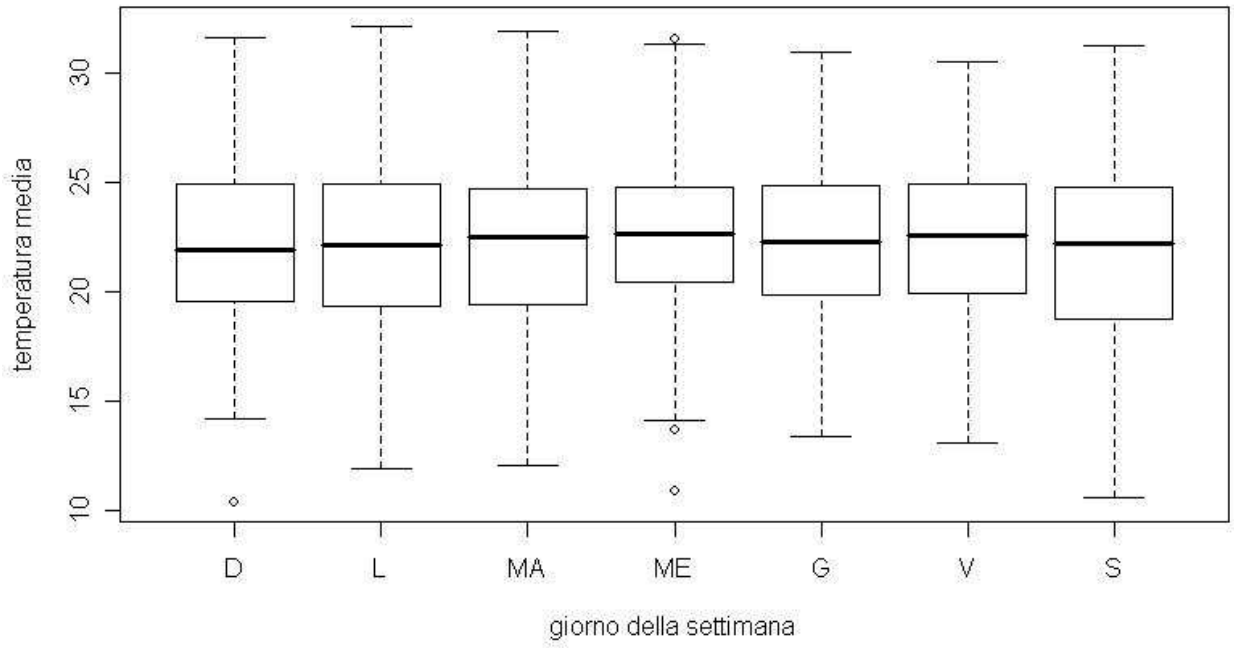

Figura 15: Diagramma a scatola della temperatura in funzione del giorno della settimana.

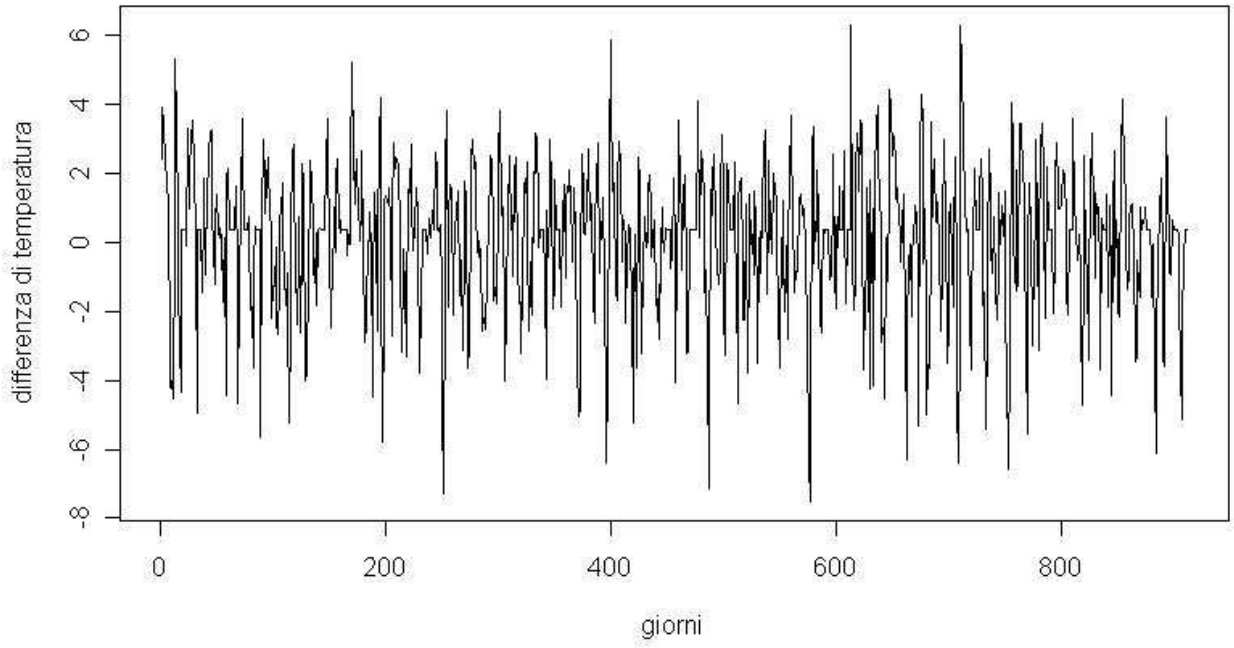

Figura 16: Andamento nel tempo della variazione di temperatura.

Per quanto riguarda la variazione di temperatura, essa è definita come la differenza tra la media giornaliera e la media dei valori registrati nei tre giorni precedenti. Presenta un valore mediano sull'intero periodo pari a 0.34 °C con un intervallo interquartile compreso tra -1.01 e 1.42 °C; osservandone l'andamento nel tempo (Figura 16) e i diagramma a scatola in funzione dell'anno (Figura 17), si può affermare che i valori di questa variabile non abbiano subito variazioni rilevanti nel corso del periodo analizzato.

Anche le Figure 18 e 19, che rappresentano i diagramma a scatola in funzione del mese e del giorno della settimana, confermano questa affermazione. Si possono comunque notare due particolarità, una riguardante gli anni e l'altra i mesi; nel 1998 e nel 2002 si osserva una variabilità leggermente superiore a quella degli altri anni (Figura 17); nei mesi di maggio ed agosto si sono avute variazioni di temperatura più frequenti, come si può desumere dai "baffi" dei diagramma a scatola relativi a questi mesi (Figura 18), che sono leggermente più lunghi di quelli degli altri mesi.

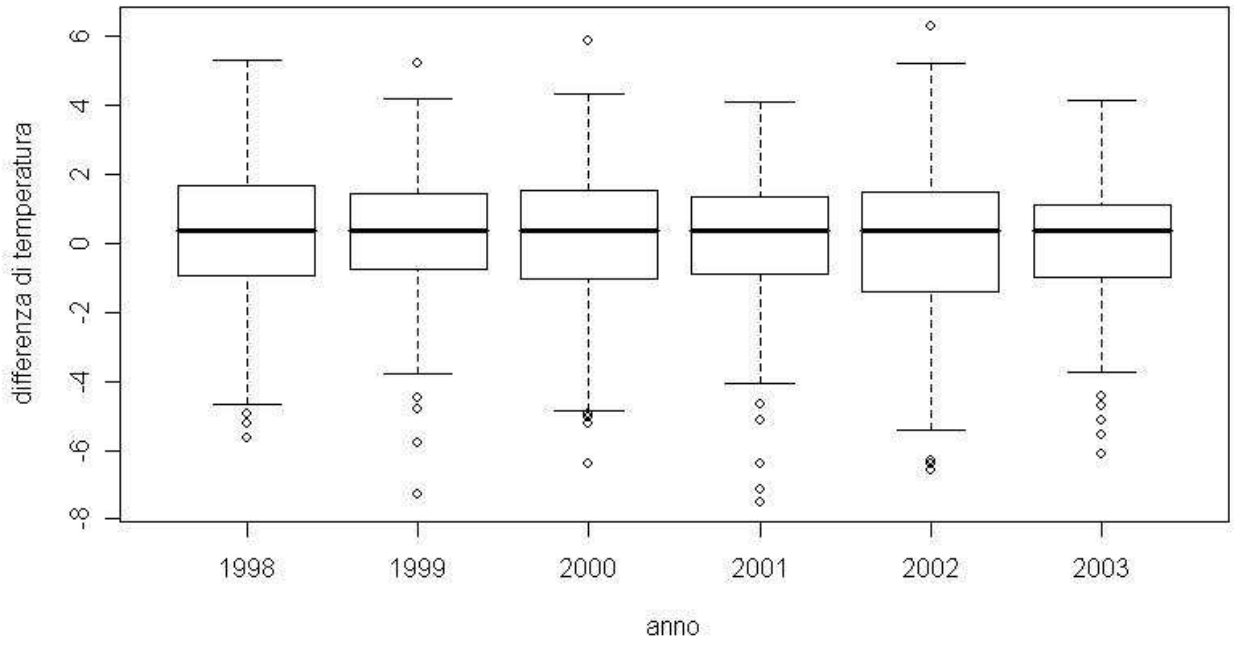

Figura 17: Diagramma a scatola della variazione di temperatura in funzione dell'anno.

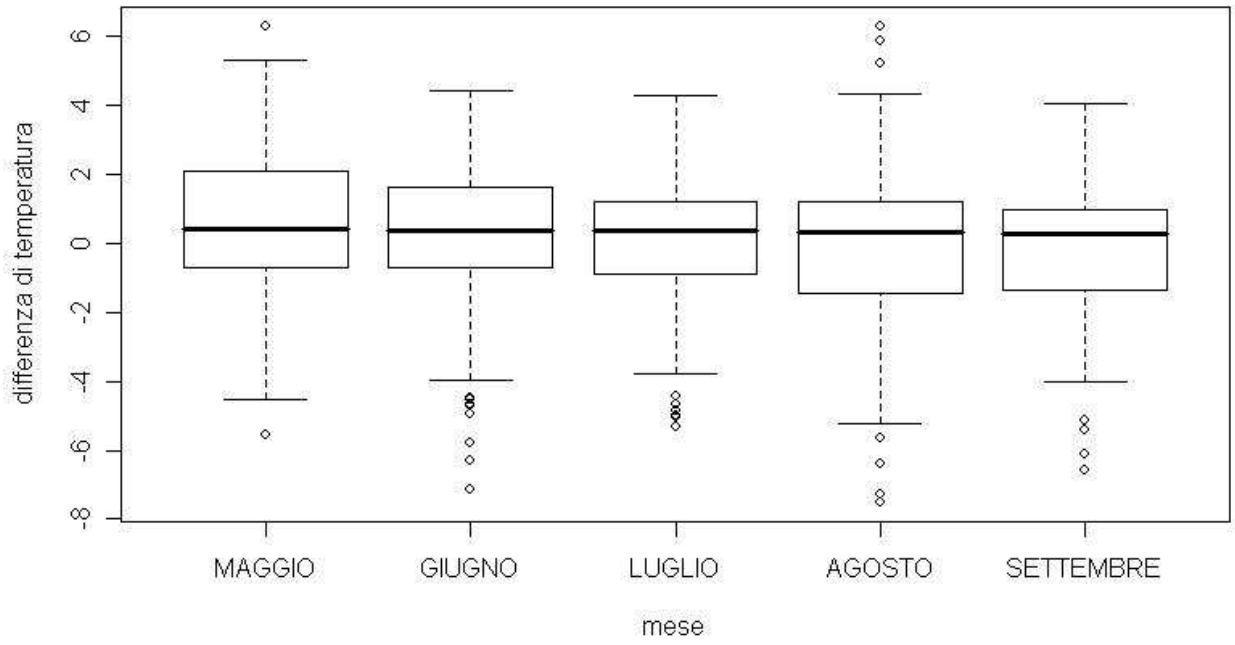

Figura 18: Diagramma a scatola della variazione di temperatura in funzione del mese.

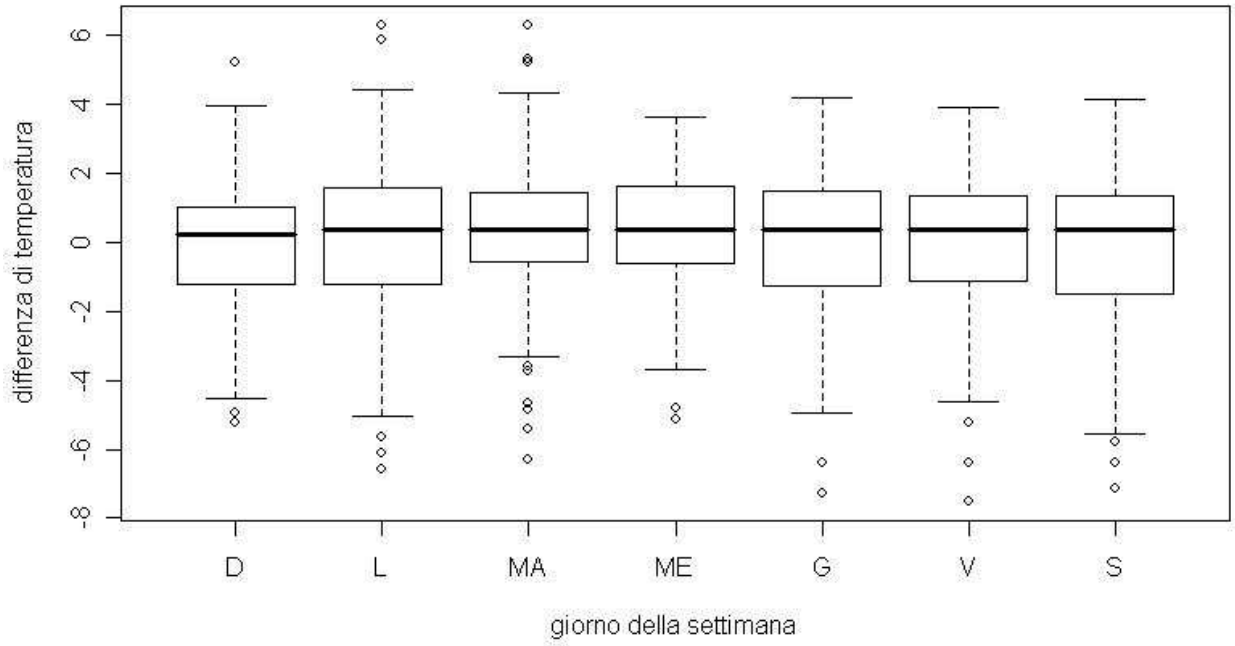

Figura 19: Diagramma a scatola della variazione di temperatura in funzione del giorno della settimana.

3.5 L'OZONO COME DATO FUNZIONALE

L'ozono e' stato trasformato in tre tipi di oggetto funzionale: con base Fourier, con base “bspline” e con base costante.

3.5.1 L'OZONO CON BASE FOURIER

La base Fourier è stata calcolata con un numero di funzioni di base pari a 5, 7, 9 e 23.

Il numero 23 è stato scelto perché è il numero di funzioni di base con cui si ottiene una base satura nel caso delle osservazioni orarie, come quelle disponibili in questa tesi per l'ozono. Ovviamente è anche la scelta più costosa in termini di tempo di calcolo, perché presenta il massimo numero di coefficienti da stimare. I numeri inferiori (5, 7 e 9) sono stati scelti allo scopo di individuare una base ridotta che offra un ragionevole bilanciamento tra qualità della trasformazione e tempo di calcolo.

Se si osserva la Figura 20 (a sinistra), si può vedere il grafico delle media funzionale dell'ozono con 23 funzioni di base; esso presenta molte oscillazioni (dovute all’alto numero di componenti) di ampiezza considerevole, che ne rendono difficile l’interpretazione. Anche il grafico della deviazione standard (Figura 20, a destra) ha questo aspetto, pertanto si può ridurre il numero delle componenti per ottenere delle rappresentazioni grafiche più facilmente interpretabili.

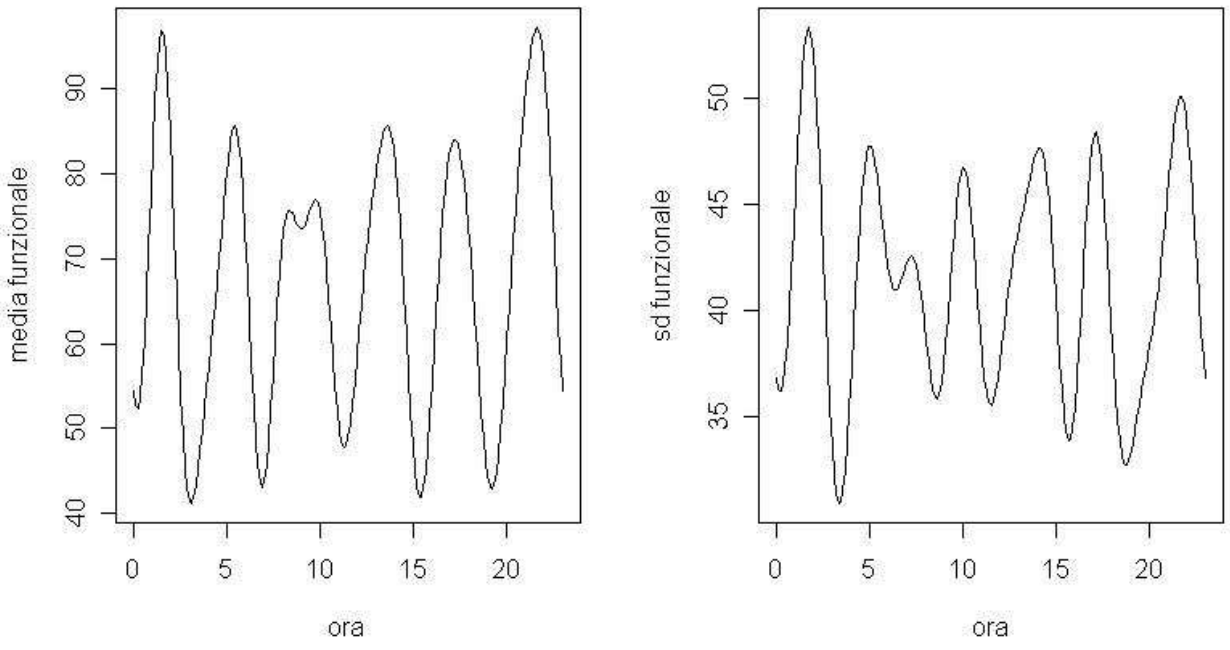

Figura 20: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base Fourier con 23 funzioni di base.

Considerando 5 funzioni di base, la media funzionale è molto più chiara ed mostra un andamento dell'ozono abbastanza coerente con le aspettative: valori in aumento in mattinata ed alti nel pomeriggio, con una diminuzione durante le ore notturne; l'unica caratteristica insolita è la crescita che si nota dopo le ore 20 (Figura 21, a sinistra).

La deviazione standard raggiunge i minimi nelle ore centrali della giornata, per poi aumentare nel pomeriggio e infine scendere durante la mattina (Figura 21 a destra). Tuttavia l'intervallo di variazione è piccolo: ha infatti lunghezza pari a 2, con la concentrazione di ozono che varia tra 31.5 e 33.5; quindi si può affermare che la variabilità sia elevata e si mantenga su valori simili.

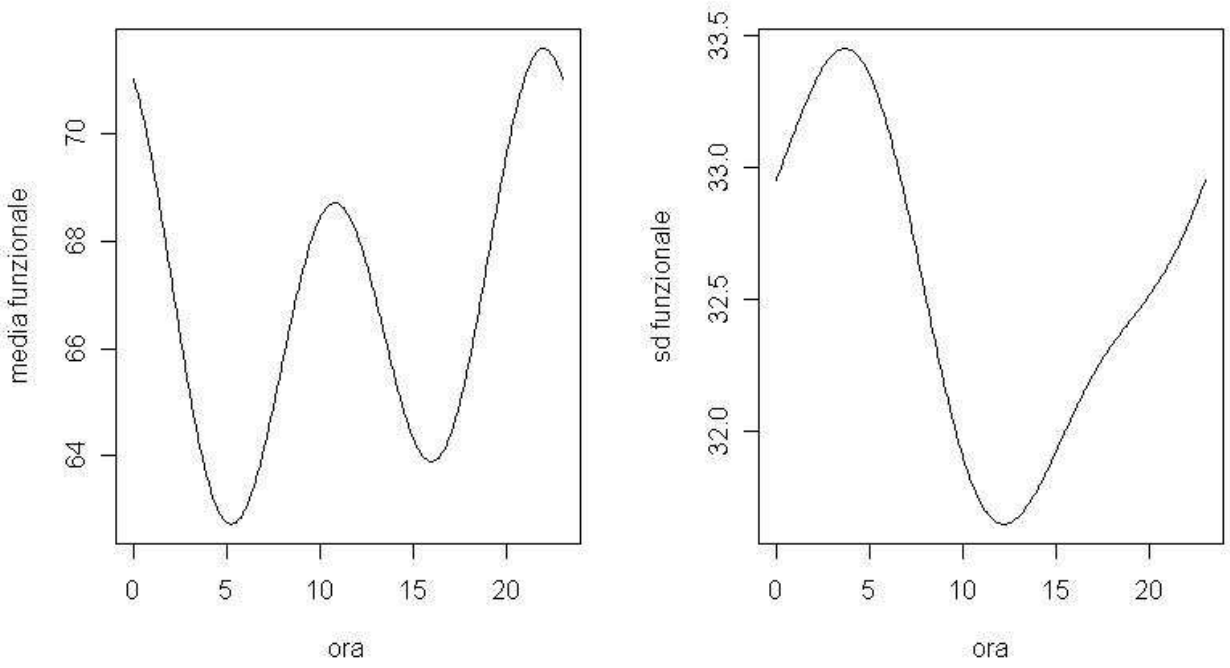

Figura 21: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base Fourier con 5 funzioni di base.

Aumentando il numero di funzioni di base a 7 e successivamente a 9, si ottengono i grafici in Figura 22 e in Figura 23 rispettivamente.

La media funzionale, pur presentando alcune differenze rispetto al caso precedente, suggerisce un'interpretazione analoga.

Nelle ore centrali del giorno, il massimo che si osservava nel caso di 5 funzioni di base (Figura 21, a sinistra) diventa doppio con 7 funzioni (Figura 22, a sinistra) e si allarga ulteriormente con 9 (Figura 23, a sinistra), aumentando così l'intervallo di tempo caratterizzato da valori elevati dell'ozono. I minimi diventano più stretti e si spostano verso gli estremi dell'intervallo 0-23, per effetto dell'aumento delle oscillazioni; infine, anche il massimo notturno diventa più stretto.

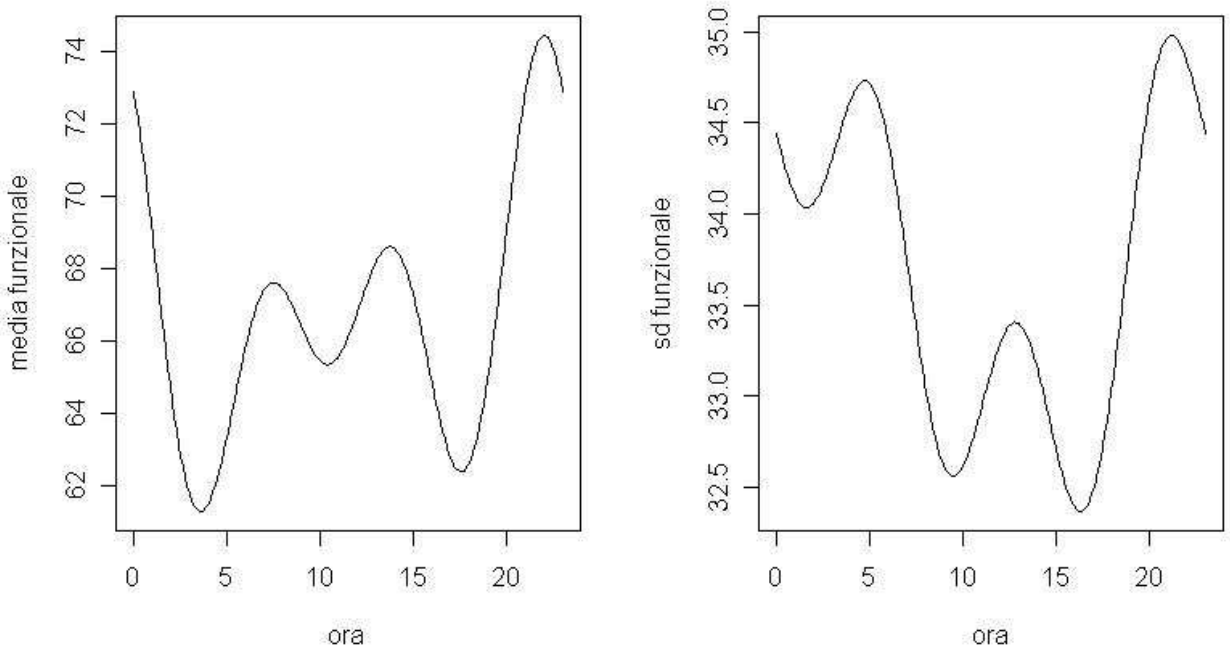

Figura 22: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base Fourier con 7 funzioni di base

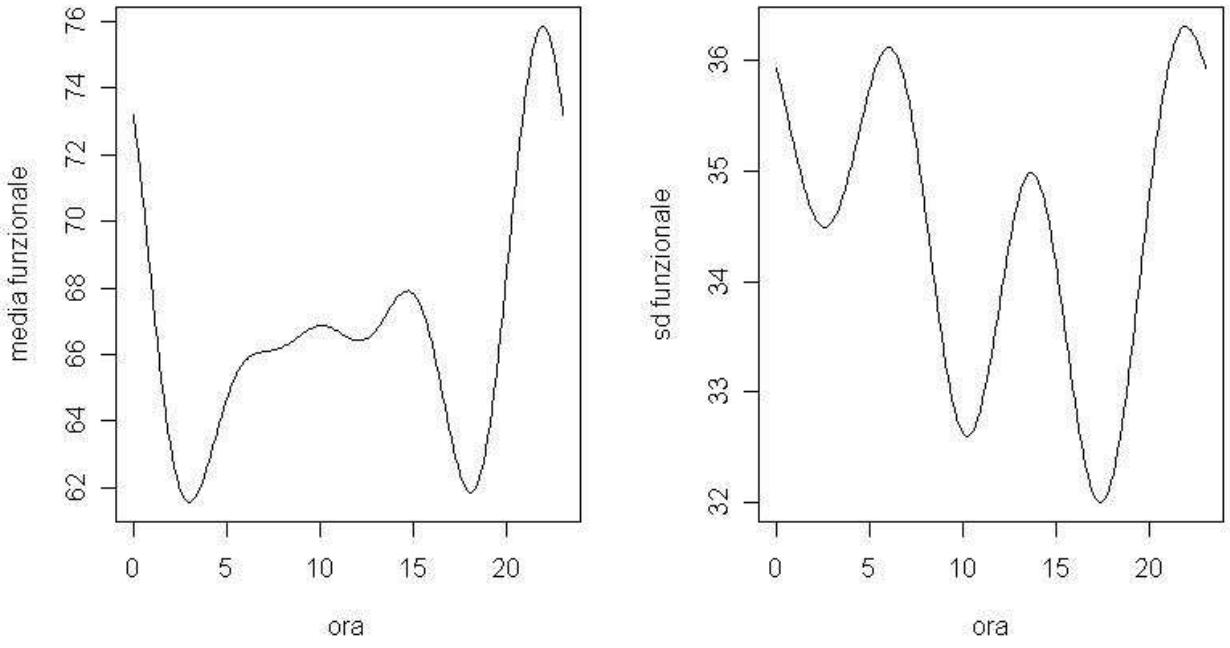

Figura 23: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base Fourier con 9 funzioni di base.

Per quanto riguarda la deviazione standard, i suoi valori aumentano passando a 7 (Figura 22, a destra) ed a 9 funzioni di base (Figura 23, a destra); inoltre nel primo pomeriggio compare un massimo, che indica una maggior variabilità della concentrazione dell'inquinante in quelle ore.
Per rendere più evidenti queste considerazioni, in Figura 24 e 25 sono rappresentate la media e la deviazione standard per i casi con 5, 7 e 9 funzioni di base insieme.

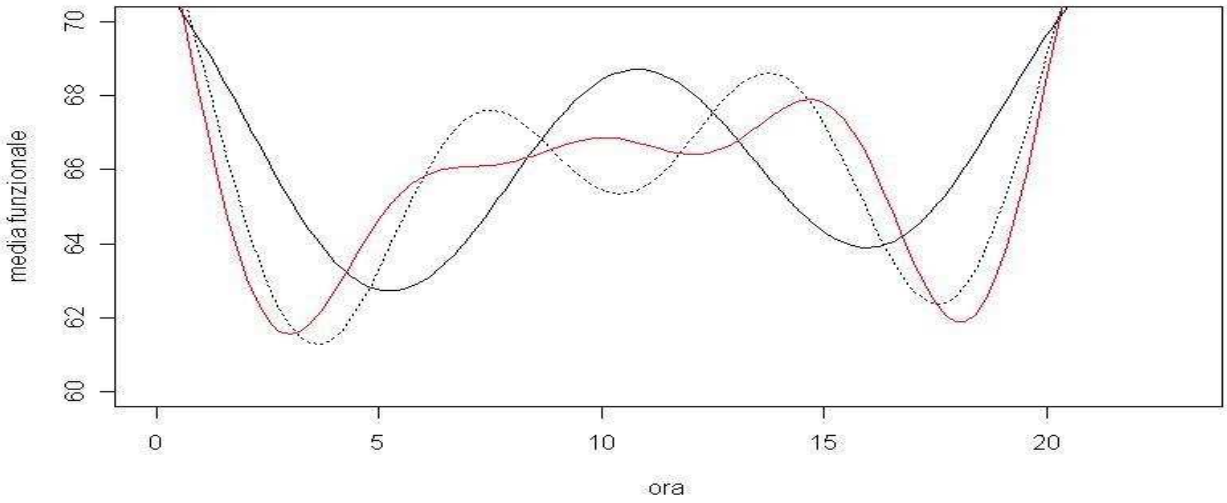

Figura 24: Media dell'ozono con base Fourier con 5 (in nero) , 7 (tratteggiato) e 9 (in grigio) funzioni di base.

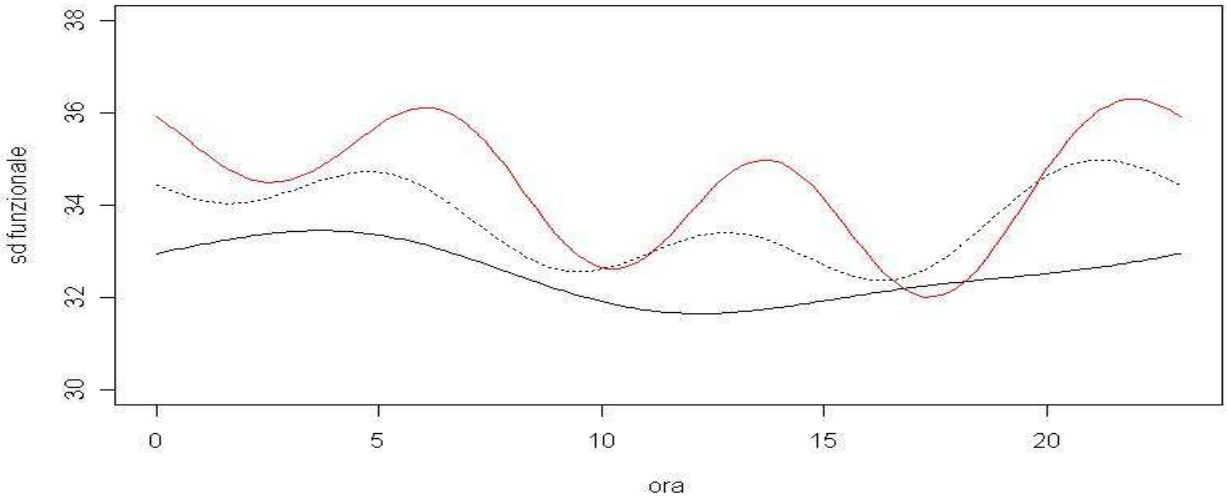

Figura 25: Deviazione standard dell'ozono con base Fourier con 5 (in nero) , 7 (tratteggiato) e 9 (in grigio) funzioni di base.

Alla luce di quanto emerso per il caso della base Fourier, è stato scelto di utilizzare nel seguito un numero di funzioni di base pari a 9 per tutte le trasformazioni, con l'obiettivo di coniugare qualità della trasformazione, variabilità non elevata e tempo di calcolo ridotto.

3.5.2 L'OZONO CON BASE BSPLINE

La base "bspline", calcolata pertanto con 9 funzioni di base, presenta valori elevati della media in corrispondenza delle ore centrali ed inoltre due picchi notturni, probabilmente espressione dell'accumulo di ozono che si verifica durante le ore di luce solare (figura 26, a sinistra).
La deviazione standard è più alta nelle ore notturne e più bassa nell'intervallo 5-17, ma nel complesso presenta dei valori elevati (Figura 26, a destra).

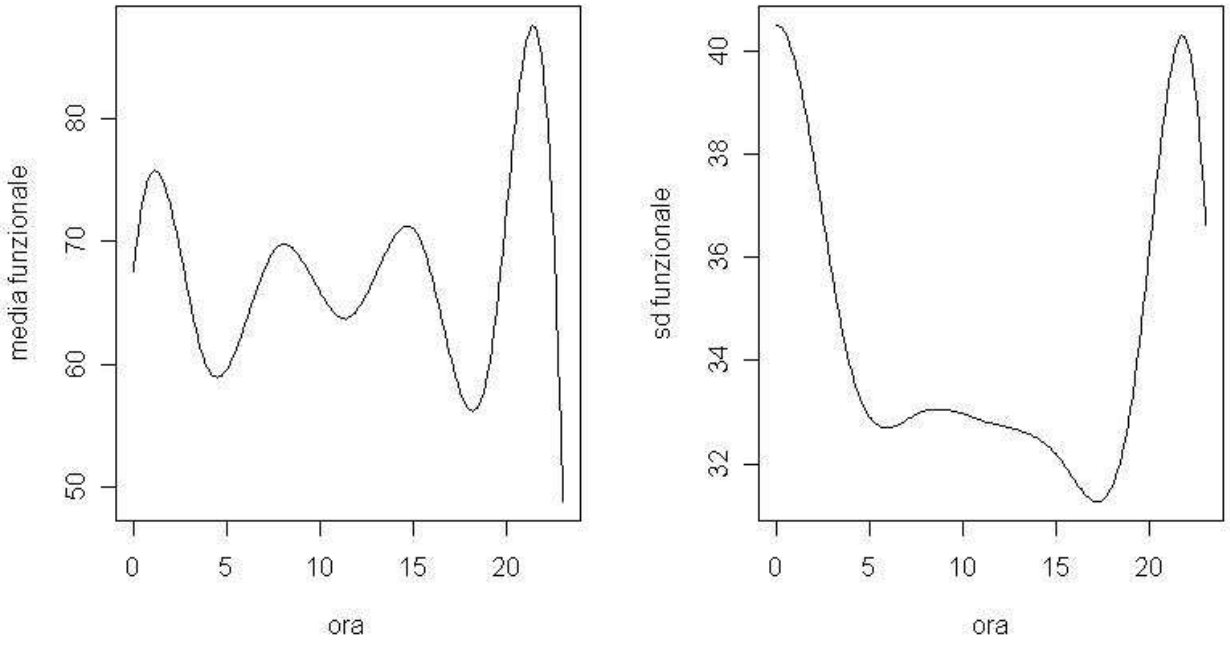

Figura 26: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base bspline con 9 funzioni di base.

Confrontando con quanto ottenuto per la base Fourier (con egual numero di funzioni di base), si nota come la media per la base bspline assuma valori in un intervallo più ampio e presenti dei massimi e dei minimi più delineati, ma sostanzialmente confermi l'andamento ottenuto con la base precedente (Figura 27, a sinistra).
La deviazione standard per la base bspline assume valori più alti nelle ore notturne, mentre durante le ore diurne mostra una variabilità inferiore rispetto alla base Fourier, che invece presenta anche due massimi in questo intervallo di tempo (Figura 27, a destra).

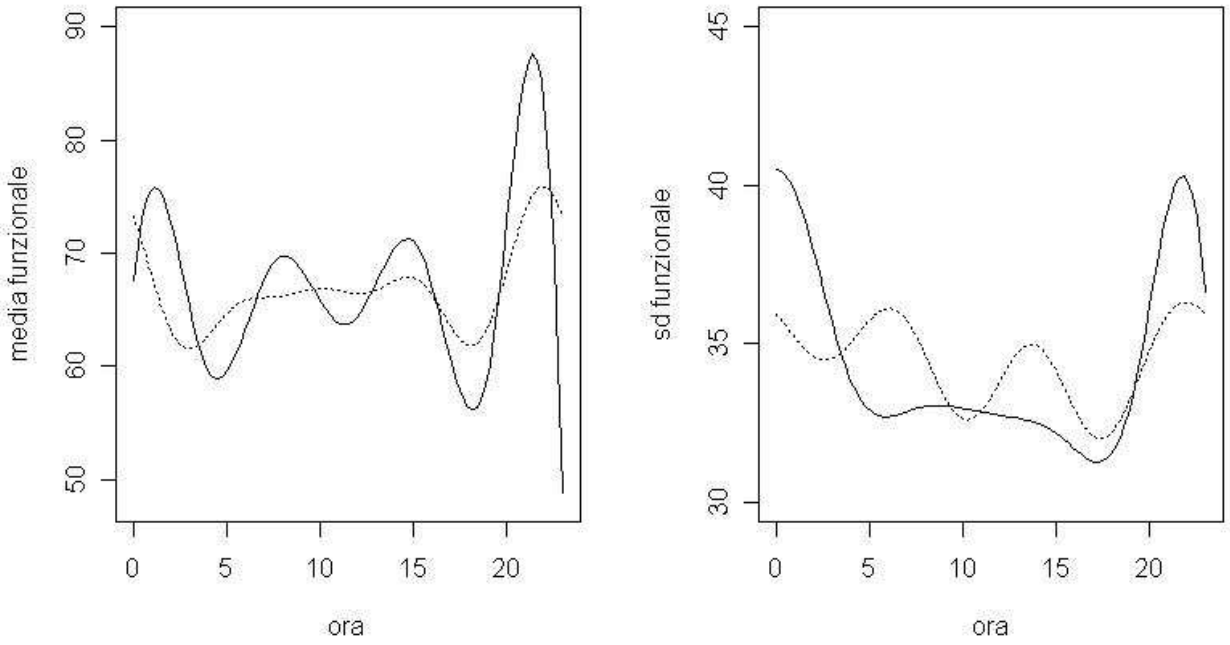

Figura 27: Media (a sinistra) e deviazione standard (a destra) per l'ozono con base bspline (in nero) e con base Fourier (tratteggiato) con 9 funzioni di base.

3.5.3 L'OZONO CON BASE COSTANTE

La base costante crea un oggetto funzionale a partire dalle rilevazioni giornaliere (non più quelle orarie) ed è stata usata in questa tesi come termine di paragone per confrontare le prestazioni ottenute con le altre due basi. Lo scopo è verificare l'effettiva necessità di utilizzare i dati orari invece dei più semplici dati giornalieri.

Di seguito è riportato il grafico della media funzionale (insieme a quelli delle basi Fourier e della base bspline con 9 funzioni basis); ovviamente, poiché il valore giornaliero viene replicato per tutte le 24 ore del giorno, il grafico risulta una retta orizzontale. Per lo stesso motivo, non è possibile produrre il grafico dell'errore standard.

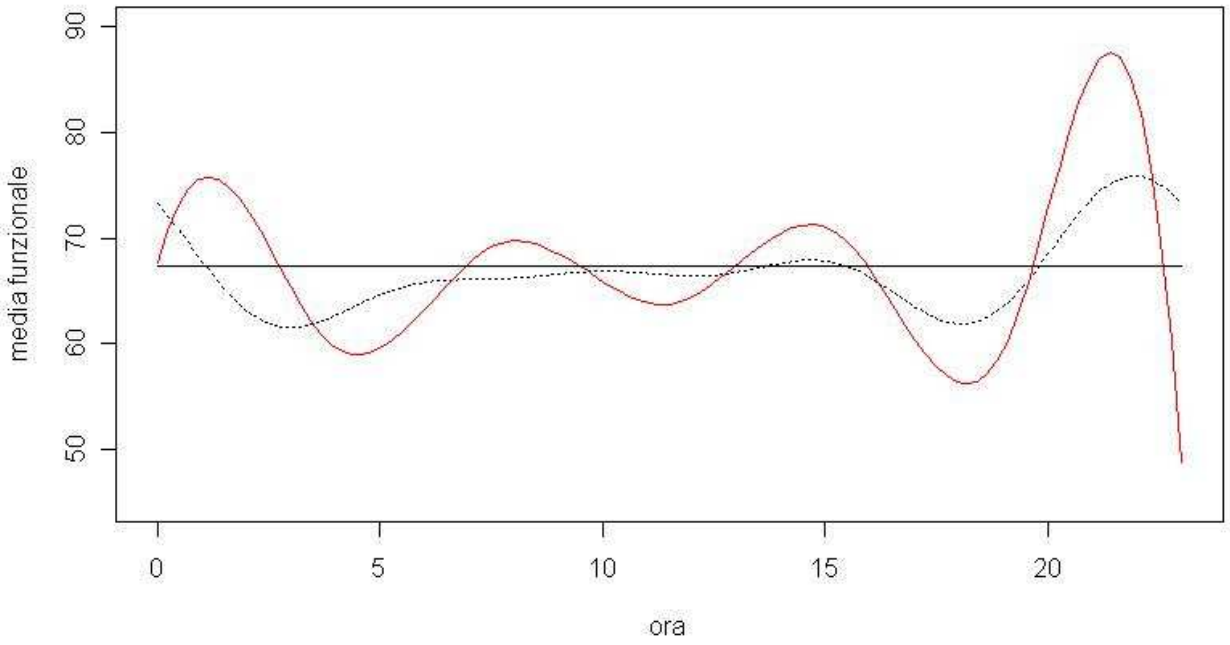

Figura 28: Media per l'ozono con base costante (in nero), con base "fourier" (tratteggiato) e con base "bspline" (in grigio).

4. I MODELLI

I modelli di regressione qui presentati sono modelli lineari calcolati con la funzione *fRegress()* della libreria FDA; la formulazione generale di questi modelli è la seguente:

$$Y_i = \beta_0 + \int \beta_1(t) x_{i1}(t)\,dt + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711 \quad t = 1,..,24$$

dove:

- i indica il giorno e t l'ora,
- Y_i rappresenta una trasformazione del numero dei ricoveri,
- $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale,
- $x_{i2},\ldots,x_{i7}$ sono le variabili confondenti,
- $\beta_1(t)$ è il coefficiente di regressione dell'ozono, espresso in forma funzionale,
- $\beta_2,\ldots,\beta_7$ sono i coefficienti di regressione delle variabili confondenti,
- ε_i è il termine di errore.

Si assume che i termini di errore siano indipendenti, abbiano media nulla ed abbiano varianza comune.

La variabile risposta è rappresentata dal numero dei ricoveri giornalieri; poiché questa è una variabile di conteggio, al fine di riportare i dati ad una condizione di normalità, nei modelli dei paragrafi seguenti sono state considerate due sue trasformazioni, ovvero:

$$y_{iA} = \log(r_i)$$

e

$$y_{iB} = \sqrt{r_i} + \sqrt{r_i + \frac{1}{2}}$$

dove r_i rappresenta il numero dei ricoveri giornalieri nel giorno i.

Tra i regressori, solo l'ozono è stato utilizzato dopo essere stato trasformato in dato funzionale, perchè è la variabile su cui si vuole focalizzare l'attenzione in questa tesi. L'ozono in forma funzionale è espresso dalla formula

$$x_{i1}(t) = \sum_{k=1}^{K} c_{ik}\,\phi_k$$

dove c_{ik} sono i coefficienti reali e ϕ_k le funzioni di base.

In modo analogo, il coefficiente di regressione dell'ozono è espresso in forma funzionale da

$$\beta_1(t) = \sum_{k=1}^{K} \beta_k\,\phi_k$$

dove β_k sono i coefficienti reali e ϕ_k le funzioni di base.

Le variabili confondenti, invece, sono state inserite nella regressione come dati non funzionali; per fare ciò è stato necessario applicare ad ognuna una base costante, perchè la funzione di R che calcola la regressione funzionale richiede i regressori in forma funzionale. Tuttavia, come già detto in precedenza, l'utilizzo di una base costante non comporta di fatto una trasformazione dei dati (come avviene nel caso di una base Fourier o bspline) ma si limita a riportare i dati discreti nella forma funzionale.

Le variabili confondenti sono state etichettate come:

- temperatura : la temperatura media giornaliera,
- tempo : l'indicatore progressivo giornaliero,
- pm10 : la concentrazione giornaliera di PM10,
- festa : l'indicatore del giorno festivo,
- Tscarto : la variazione di temperatura,
- Wday : l'indicatore del giorno della settimana.

Nei paragrafi seguenti sono presentati i modelli di regressione studiati. Per ognuno sono state calcolate le stime dei coefficienti, degli errori standard, le statistiche t e la statistica R^2 ; le statistiche t significative al 5% sono state evidenziate tramite sottolineatura. Viene infine presentato il grafico della risposta stimata (in grigio) sovrapposto a quello della variabile risposta (in nero).

4.1 VARIABILE RISPOSTA y_{IA} E OZONO CON BASE FOURIER.

L'equazione di questo modello è la seguente:

$$y_{iA} = \beta_0 + \beta_1 t + \int x_{i1}(t)\,dt + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711 \quad t = 1,..,24$$

dove $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale con una base di Fourier con 9 funzioni di base.

Nella tabella seguente vengono presentate le stime ottenute:

	Coefficiente	*Deviazione standard*	*Statistica t*
costante	0.1538671	0.004076673	37.74331
ozono	2.027945e-04 4.848330e-05 1.414090e-06 -1.340972e-04 -2.705971e-04 -1.985696e-04 -7.471294e-04 -3.465612e-04 3.477018e-04	5.989106e-04 -3.053998e-05 -6.206656e-06 -1.075041e-05 -4.504611e-06 3.687716e-06 -6.832488e-06 -8.817731e-06 -5.935162e-06	0.3386056 -1.5875352 -0.2278345 12.4736754 60.0711320 -53.8462349 109.3495399 39.3027654 -58.5833696
temperatura	-0.0004211225	8.253992e-05	-5.102047
tempo	0.009936	0.001527916	6.502977
pm10	0.0001622447	2.157462e-05	7.520162
festa	0.0003087728	0.002396628	0.1288364
Tscarto	0.0002767458	0.0001392761	1.98703
wday	-0.001263516	0.0001374857	-9.190157
Rsqrd=0.1915520			

Tranne l'indicatore del giorno festivo, tutti gli altri regressori risultano significativi; in particolare, sei dei nove coefficienti delle funzioni di base dell'ozono sono significativi. Questo sottolinea l'importanza dell'azione di questo inquinante atmosferico sulla salute umana.

In Figura 29 è presentato il grafico della risposta stimata (in grigio) e quello della variabile risposta (in nero).

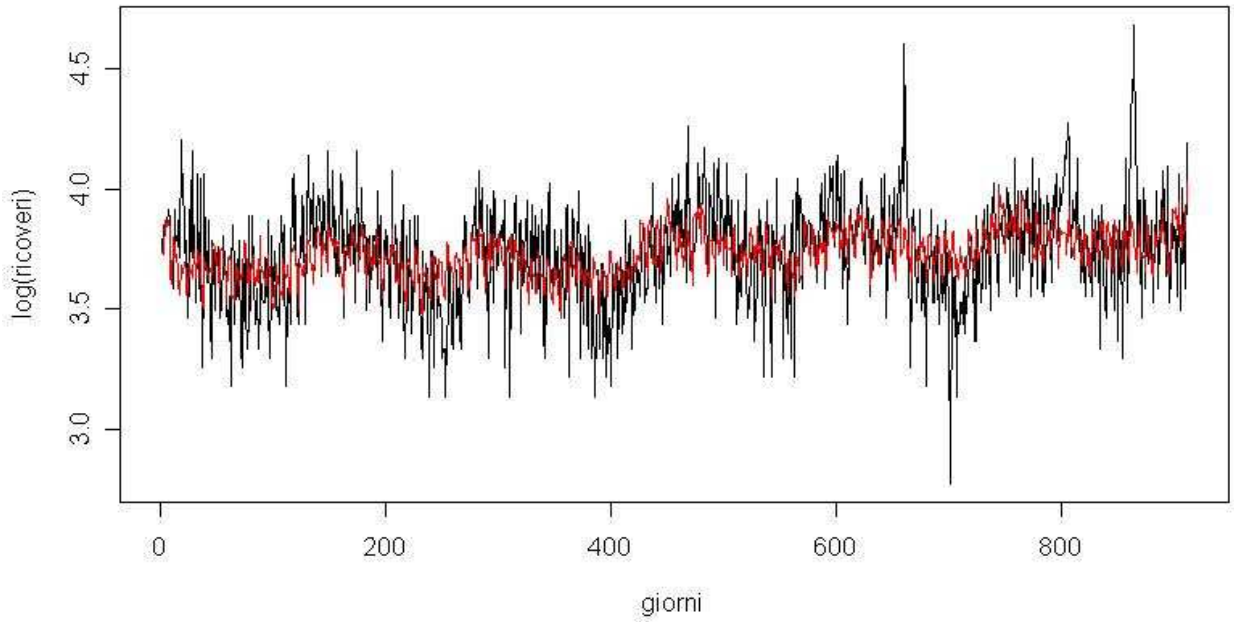

Figura 29: Grafico della variabile risposta (in nero) e grafico della risposta stimata (in grigio) in funzione del tempo.

4.2 VARIABILE RISPOSTA y_{iB} E OZONO CON BASE FOURIER.

L'equazione di questo modello è la seguente:

$$y_{iB} = \beta_0 + \beta_1 \int x_{i1}(t)\,dt + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711 \quad t = 1,..,24$$

dove $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale con una base di Fourier con 9 funzioni di base.

Nella tabella seguente vengono presentate le stime ottenute:

	Coefficiente	*Deviazione standard*	*Statistica t*
costante	0.5202766	0.1942307	2.678653
ozono	1.676625e-03 2.991888e-04 -1.644473e-05 -1.149775e-03 -1.764387e-03 -1.440699e-03 -4.478809e-03 -2.162091e-03 2.027680e-03	0.0285347468 -0.0014550596 -0.0002957125 -0.0005121972 -0.0002146196 0.0001756991 -0.0003255299 -0.0004201156 -0.0002827773	0.05875731 -0.20561961 0.05561052 2.24479049 8.22099647 -8.19981091 13.75851912 5.14641797 -7.17059109
temperatura	-0.002453374	0.003932566	-0.623861
tempo	0.06580433	0.07279665	0.9039472
pm10	0.001054063	0.001027910	1.025442
festa	4.736364e-05	0.1141859	0.0004147941
Tscarto	0.001599614	0.006635729	0.2410608
wday	-0.0081109	0.006550428	-1.238224
Rsqrd=0.1910912			

In questo caso, nonostante un R^2 simile a quanto ottenuto precedentemente, la significatività dei coefficienti si riduce a metà delle funzioni di base dell'ozono ed a nessuna tra le variabili confondenti. I valori della variabile risposta così modificata sembrano quindi influenzati solo dall'ozono.

Pertanto, nel seguito sarà utilizzata per la regressione solo la prima trasformazione del numero di ricoveri, ovvero $y_{iA} = \log(r_i)$.

In Figura 30 è presentato il grafico della risposta stimata (in grigio) e quello della variabile risposta (in nero).

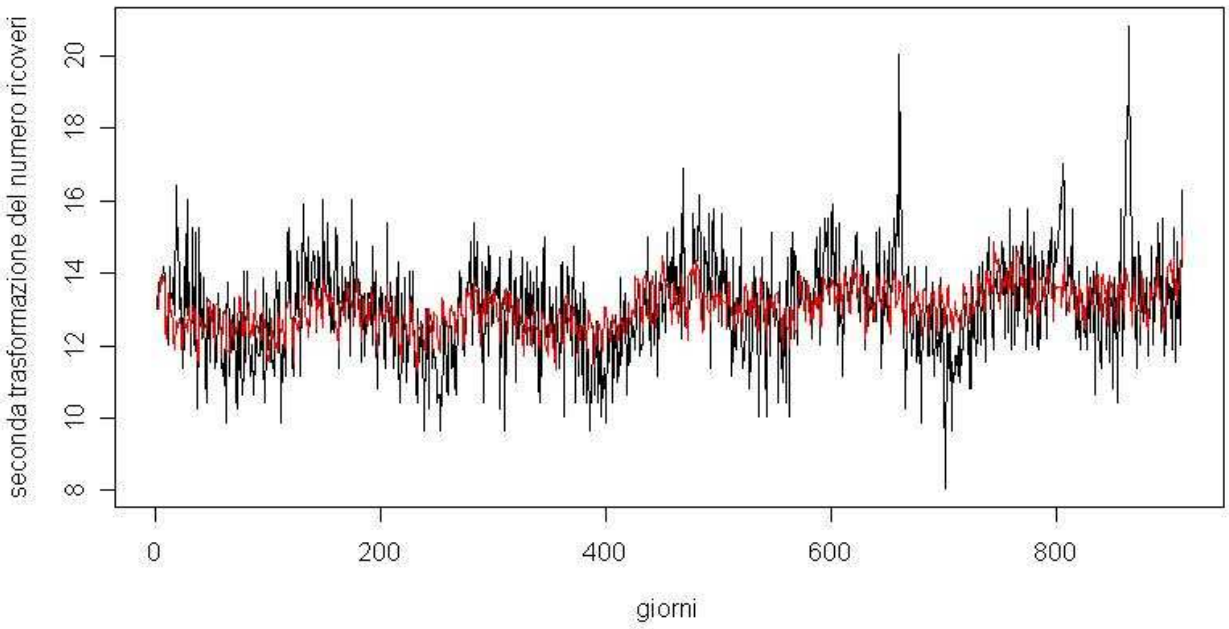

Figura 30: Grafico della variabile risposta (in nero) e grafico della risposta stimata (in grigio) in funzione del tempo.

4.3 VARIABILE RISPOSTA y_{iA} E OZONO CON BASE BSPLINE.

L'equazione di questo modello è la seguente:

$$y_{iA} = \beta_0 + \beta_1 \int x_{i1}(t)\,dt + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711 \quad t = 1,..,24$$

dove $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale con una base bspline con 9 funzioni di base.

Nella tabella seguente vengono presentate le stime ottenute:

	Coefficiente	*Deviazione standard*	*Statistica t*
costante	0.1610838	0.004260987	37.80435
ozono	-7.746725e-04 2.584624e-04 5.677160e-04 -4.914437e-04 5.811338e-04 -6.539208e-04 8.023700e-04 -3.045969e-05 -5.219496e-04	4.311580e-04 3.878862e-05 1.798926e-04 6.518537e-05 1.209496e-04 6.393081e-05 1.911577e-04 3.889805e-05 4.374224e-04	-1.7967256 6.6633571 3.1558611 -7.5391717 4.8047587 -10.2285708 4.1974254 -0.7830647 -1.1932392
temperatura	-0.000460769	8.591682e-05	-5.362966
tempo	0.01044515	0.001614705	6.468767
pm10	0.0001711331	2.259685e-05	7.573318
festa	0.0003185926	0.002503193	0.1272745
Tscarto	0.0002899816	0.0001455253	1.992654
wday	-0.001317461	0.0001437932	-9.16219
Rsqrd=0.187681			

Come per la regressione del paragrafo 4.1 , le variabili confondenti sono tutte significative a parte l'indicatore del giorno festivo; anche i valori delle statistiche t sono molto simili.

L'ozono con base bspline risulta significativo e, come ottenuto con la base Fourier, lo sono sei dei suoi coefficienti.

Il valore del R^2 (pari a 0.187681) è più basso, ma non si discosta in modo apprezzabile dai valori delle regressioni precedenti.

In Figura 31 è presentato il grafico della risposta stimata (in grigio) e quello della variabile risposta (in nero).

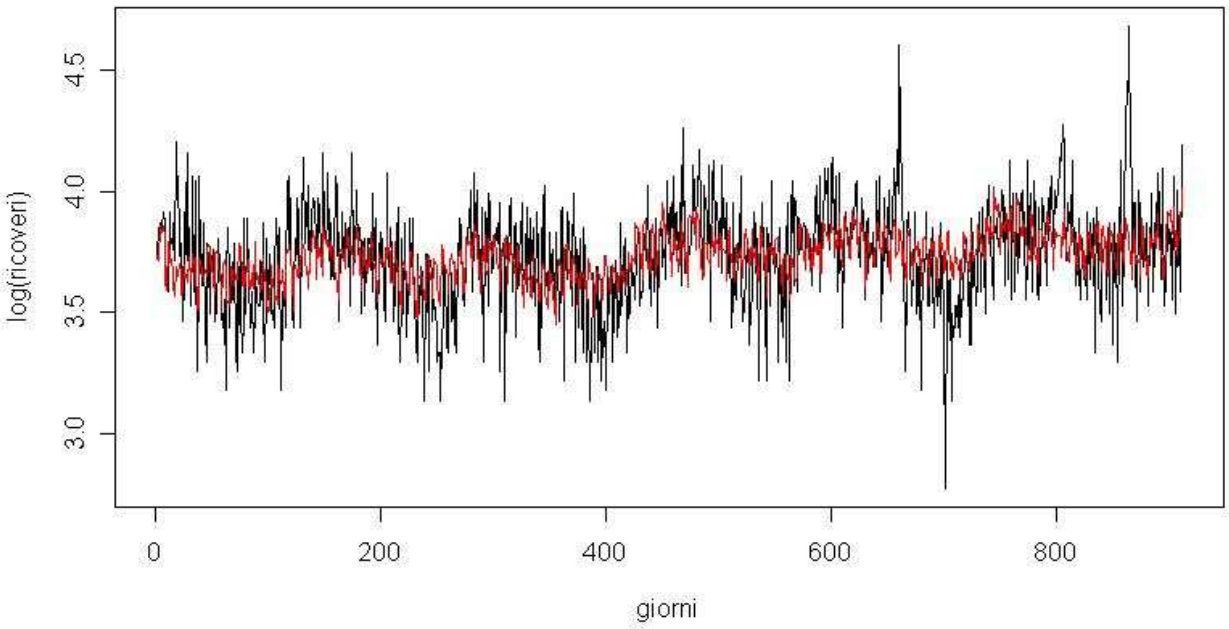

Figura 31: Grafico della variabile risposta (in nero) e grafico della risposta stimata (in grigio) in funzione del tempo.

4.4 VARIABILE RISPOSTA y_{iA} E OZONO CON BASE COSTANTE.

L'equazione di questo modello è la seguente:

$$y_{iA} = \beta_0 + \beta_1 \int x_{i1}(t)\,dt + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711 \quad t = 1,..,24$$

dove $x_{i1}(t)$ rappresenta l'ozono considerato come dato funzionale con una base costante.

E' opportuno specificare che in questo caso vengono considerate le rilevazioni giornaliere dell'ozono e non quelle orarie; inoltre si ricorda che l'applicazione della base costante significa che nella regressione sono usati i valori giornalieri reali dell'ozono; per questo motivo, a differenza delle tabelle precedenti, in quella seguente il coefficiente dell'ozono è uno solo.

Questo modello è stato costruito per indagare sull'effettiva utilità delle rilevazioni orarie invece dei più semplici dati giornalieri.

Nella tabella seguente vengono presentate le stime ottenute:

	Coefficiente	*Deviazione standard*	*Statistica t*
costante	0.1629742	0.002143815	76.02065
ozono	1.153419e-05	1.818343e-05	0.6343239
temperatura	-0.0005071691	0.0001263874	-4.012813
tempo	0.01124309	0.001523604	7.379275
pm10	0.0001788431	2.294986e-05	7.792776
festa	-0.0002766765	0.002531198	-0.1093066
Tscarto	0.0002646761	0.0001477107	1.791855
wday	-0.001322715	0.0001453231	-9.101892
Rsqrd=0.1682723			

Come si può vedere nella tabella, in questa forma l'ozono non risulta significativo (il valore della t è circa 0,63); ciò significa che l'influenza dell'ozono sui ricoveri si manifesta solo quando si considerano le informazioni contenute nelle rilevazioni orarie, per cui si può affermare che l'andamento della concentrazione durante le ore della giornata sia più importante della relativa quantità media giornaliera. Infatti, la conoscenza dell'andamento orario permette di capire se la concentrazione abbia superato una certa soglia di allarme e di valutare la durata della permanenza oltre questa soglia.

Per quanto riguarda le variabili confondenti, si conferma il fatto che i giorni festivi non incidono sui ricoveri; in aggiunta, in questa regressione anche la differenza di temperatura rispetto ai giorni precedenti (rappresentata da Tscarto) non risulta significativa; le altre variabili sono invece tutte significative.

Il valore del R^2 (pari a 0.1682723) indica una percentuale di variabilità spiegata intorno al 17% , non molto diversa dal 18-19% delle regressioni precedenti.

In Figura 32 è presentato il grafico della risposta stimata (in grigio) e quello della variabile risposta (in nero).

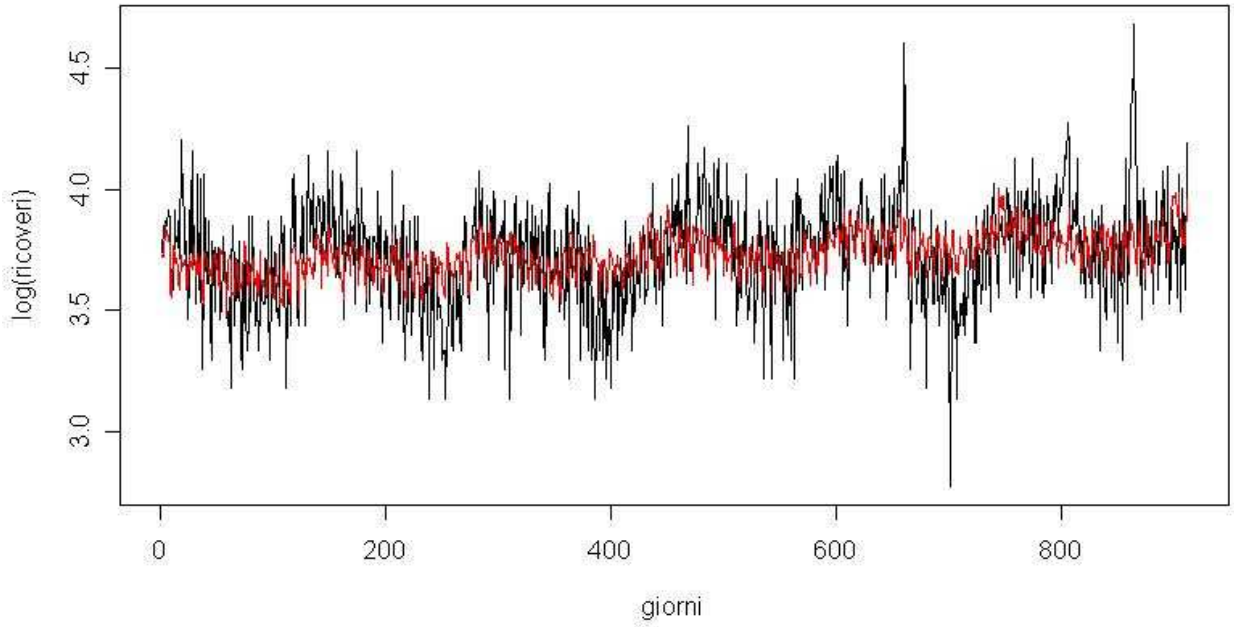

Figura 32: Grafico della variabile risposta (in nero) e grafico della risposta stimata (in grigio) in funzione del tempo.

4.5 MODELLO LINEARE CON RISPOSTA y_{iA}.

In questo paragrafo, viene presentato il modello lineare analogo al modello funzionale precedente. L'obiettivo è verificare se un modello lineare ed un modello funzionale con base costante producano gli stessi risultati.

L'equazione di questo modello è la seguente:

$$y_{iA} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \varepsilon_i$$

$$i = 1,\ldots,711$$

dove x_{i1} rappresenta la concentrazione media dell'ozono nel giorno i.

Nella tabella seguente vengono presentate le stime ottenute:

	Coefficiente	*Deviazione standard*	*Statistica t*
costante	3.7484071	0.0495257	75.686
ozono	0.0002653	0.0004201	0.632
temperatura	-0.0116649	0.0029198	-3.995
tempo	0.2585912	0.0351978	7.347
pm10	0.0041134	0.0005302	7.758
festa	-0.0063636	0.0584749	-0.109
Tscarto	0.0060875	0.0034124	1.784
wday	-0.0304224	0.0033572	-9.062
Rsqrd=0.1683			

Confrontando questa tabella con quella del modello precedente, si nota che nel caso del modello lineare i valori dei coefficienti e degli errori standard sono più grandi; tuttavia essi mantengono lo stesso segno dei valori corrispondenti per il modello precedente ed inoltre l'aumento degli errori standard è proporzionato a quello dei coefficienti, per cui i valori delle statistiche t sono quasi identici. Pertanto si può dire che le conclusioni suggerite dai due modelli sono le stesse.

In Figura 33 sono rappresentati in ascissa i valori stimati dal modello con base costante e in ordinata quelli stimati dal modello lineare: le coppie dei punti si dispongono lungo una retta passante per il punto (0,0) e di pendenza 45°, il che mostra come i due modelli forniscano stime identiche per ogni giorno.

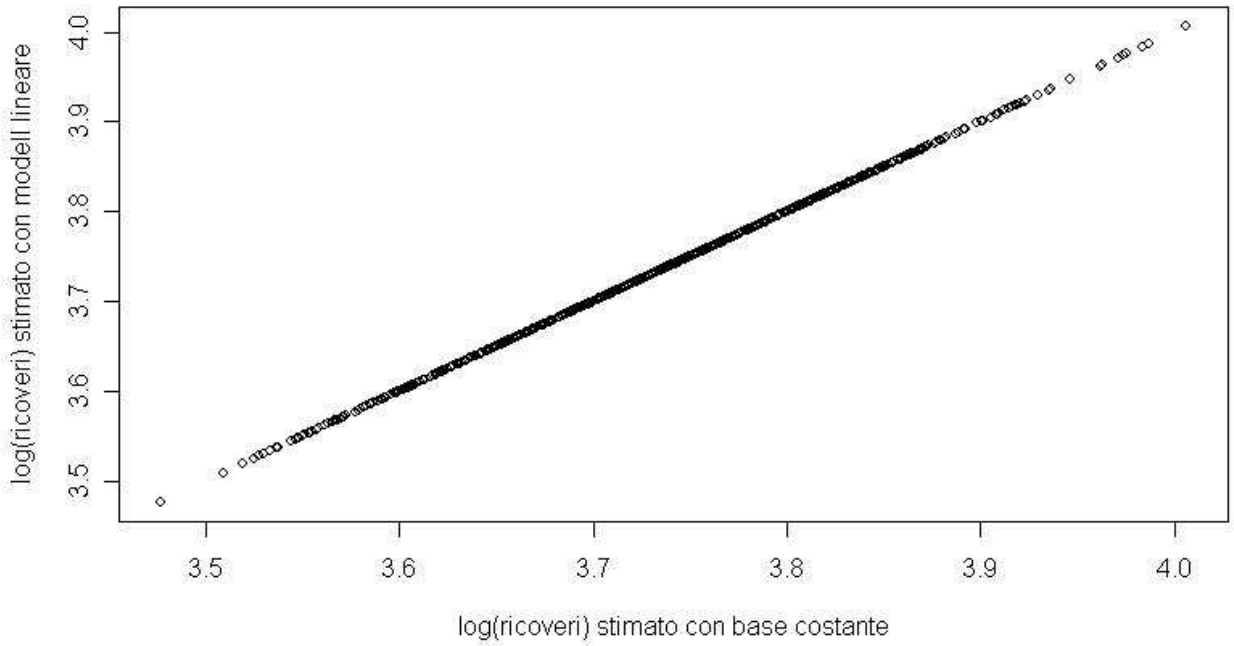

Figura 33: Grafico della risposta stimata con base costante (in ascissa) contro la risposta stimata con il modello lineare semplice (in ordinata).

5. LE CONCLUSIONI.

L'obiettivo di questa tesi era studiare la relazione tra l'ozono e la salute umana, utilizzando l'analisi dei dati funzionali.

Nel caso esaminato della città di Milano, relativamente al periodo dal 1998 al 2003, si può affermare che l'ozono abbia svolto un ruolo significativo come inquinante atmosferico; il suo effetto è stato misurato tramite la relazione con i ricoveri ospedalieri avvenuti in città in quel periodo.

Ma le analisi effettuate hanno portato anche ad altre osservazioni.

Innanzitutto, è stata constatata l'importanza di disporre delle rilevazioni orarie dell'ozono invece delle più semplici rilevazioni medie giornaliere: infatti, utilizzando queste ultime nella regressione, l'ozono non risultava significativo (statistica t pari a 0,72). La conoscenza dell'andamento orario permette di capire se la concentrazione abbia superato una certa soglia di allarme e di valutare la durata della permanenza oltre questa soglia: pertanto, in accordo con la letteratura medica, si può confermare che l'ozono sia pericoloso per la salute solo quando si mantiene ad elevate concentrazioni per un certo intervallo di tempo.

Per poter esprimere in modo completo l'informazione fornita dalle rilevazioni orarie, si è rivelata molto utile l'analisi dei dati funzionali: come già detto in precedenza, essa ha permesso di esprimere la complessità delle informazioni sull'ozono e di verificare la relazione esistente tra questo inquinante atmosferico e la salute della popolazione cittadina.

Da un punto di vista strettamente tecnico, si è potuto osservare come sia la base Fourier sia quella bspline esprimessero in modo adeguato le caratteristiche dell'ozono. Inoltre, la scelta di un numero contenuto di funzioni di base (in questo caso 9) si è rivelata più conveniente dell'utilizzo di una base satura (con 23 funzioni di base, sempre riferendosi al caso specifico qui trattato): è stato ottenuto un minor costo dal punto di vista computazionale e una maggior chiarezza nell'interpretazione delle curve, favorita dalla presenza di un numero inferiore di oscillazioni.

Per quanto riguarda le variabili confondenti, queste si sono rivelate tutte significative tranne l'indicatore del giorno festivo: questo porta alla conclusione che nei giorni festivi non ci sia una variazione di ricoveri rilevante.

Al contrario, la significatività dell'indicatore del giorno della settimana suggerisce che le differenze nelle attività umane, che hanno modalità diverse nei vari giorni della settimana a seconda della posizione sociale, del lavoro e delle abitudini, abbiano un'influenza sulla salute della popolazione da non sottovalutare.

La significatività della temperatura è in accordo con quanto espresso in letteratura, dove questa variabile ha un effetto rilevante sulla salute.

Un altra conferma di quanto previsto dalla letteratura è il ruolo delle polveri sottili (pm10), un altro inquinante atmosferico molto presente nelle città.
Infine, la significatività dell'indicatore temporale progressivo indica la presenza di un trend del numero dei ricoveri nel tempo.

Riferimenti bibliografici.

R.Michelin e A. Munari (2000). “Fondamenti Di Chimica”. Cedam.

J. Ramsay e B.W.Silverman (2002). “Applied Functional Data Analysis”. Springer.

J. Ramsay e B.W.Silverman (2002). “ Functional Data Analysis”. Springer.

H.Cardot e P.Sarda. “Estimation in generalized linear models for functional data via penalized likehood”. Journal of Multivariate Analysis 92 (2005) 24-41.

M.Chiogna e P.Bellini. “Alternative air pollution measures for detecting short-term health effects in epidemiological studies”. Environmetrics 13 (2002) 55-69.

M.Escabias, A.M. Aguilera e M.J. Valderrama. “Modelling climatological data by functional logistic regression”. University of Granada.

M. J. Gareth. “Generalized linear models with functional predictors”. J.R.Statist.Soc. B (2002) 64, Part3, 411-432.

R Development Core Team (2007). “R: A language and environment for statistical computing”. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

C.P. Weisel, R.P. Cody e P.J. Lioy. “Relationship between summertime ambient ozone levels and Emergency Department visits for asthma in central New Jersey”. Environmental Health Perspective 103 (1995), Supplement 2, 97-102.

M.L. Bell, R.D. Peng e F. Dominici. “The esposure-response curve for ozone and risk of mortality and the adequancy of current ozone regulations”. Environmental Health Perspective 114 (2006), 532-536.

APPENDICE A: codice R utilizzato.

Nota preliminare:

La versione utilizzata in questa tesi è la 2.5.1 ; su versioni precedenti alla 2.5, la libreria FDA necessaria per l'analisi dei dati funzionali non funziona.

Dataset originali:

I dataset originali sono “inq_mi_mean9203” e “dati1”.

Il primo contiene le rilevazioni orarie dell'ozono e di altri inquinanti, nonché le indicazioni temporali a cui si riferiscono le quantità rilevate (anno, mese, giorno, ora). Il periodo di riferimento va dal 1992 al 2005.

Il secondo contiene le rilevazioni giornaliere dei ricoveri ospedalieri e di altre variabili; anche qui è presente una variabile indicatrice del giorno a cui si riferisce la rilevazione. Il periodo di riferimento va dal 30 aprile 1995 al 29 settembre 2003.

Modifiche preliminari al dataset “inq_mi_mean9203”:

```
attach(inq_mi_mean9203)
# modifica della variabile “mese” (che è compresa tra 0 e 1 ) affinchè sia compresa tra 1 e 12:
mese1<-mese+1
# modifica della variabile anno per ottenere il consueto formato a 4 cifre:
anno1<-anno+1900
x<-paste(anno1,mese1,giorno)
# creazione variabile “data”:
data<-strptime(x,"%Y%m%d")
xx<-data.frame(inq_mi_mean9203,data)

# selezione rilevazioni dal 1998 al 2003, da maggio a settembre:
xxx<-subset(xx,anno1>=1998 & anno1<=2003)
x4<-subset(xxx,mese>=4 & mese<=8)
x5<-x4[145:22008,]
detach()
mese2<-mese+1

# creazione della variabile “cod”, che identifica il giorno della rilevazione tramite la data:
cod<-x5$anno*10000+mese2*100+x5$giorno
```

```
# dataset finale:
finale2<-data.frame(x5,cod)
```

Modifiche preliminari al dataset “dati1”:

```
attach(dati1)
# selezione delle variabili di interesse all'interno del dataset originale:
sottodati1<-data.frame(n ,temmax,pm10.AVE,o3.AVE,so2.AVE,no2.AVE,co.AVE,
Tscarto,tempo,wday,year,tempo, mon,festa,anno,mese,giorno,temmean,data)
detach()

# selezione rilevazioni dal 1998 al 2003, da maggio a settembre:
sottodati1<-sottodati1[467:1377,]
attach(sottodati1)

# creazione della variabile “cod”, che identifica il giorno della rilevazione tramite la data:
anno1<-anno-1900
cod<-anno1*10000+mese*100+giorno

# dataset finale:
finale1<-data.frame(sottodati1,cod)
detach()
```

Creazione della variabile risposta:

```
# numero di ricoveri giornalieri:
pz<-finale1$n

# prima trasformazione:
pzlog<-log(pz)

# seconda trasformazione:
pzmodif<-sqrt(pz)+sqrt(pz+0.5)
```

Creazione della matrice ozono a partire dal vettore medesimo:

```
o3m<-matrix(finale2$o3,byrow=T,nrow=24)
```

Creazione della variabile indicatrice del giorno di festa:

```
finale1$festa01<-ifelse(finale1$festa==T,1,0)
```

Gestione dei valori mancanti, a cui vengono sostituiti i rispettivi valori mediani:

```
# ozono:
o3m<-ifelse(is.na(o3m),27.35,o3m)
# temperatura:
finale1$temmean<-ifelse(is.na(temmean),22.14,temmean)
# pm10:
finale1$pm10.AVE<-ifelse(is.na(finale1$pm10.AVE),33.61,finale1$pm10.AVE)
```

Creazione oggetti “functional data” di diverso tipo per la variabile ozono:

```
daytime<-c(0:23)
dayrange<-c(0,23)
dayperiod <- 24

#  basis di tipo “bspline” con numero di basis pari a 9:
nbasis<-9
daybasis<-create.bspline.basis(dayrange,nbasis)
o3fdbspline<-data2fd(o3m,daytime,daybasis)

# basis di tipo costante:
# (questa basis non richiede le rilevazioni orarie, per cui viene qui utilizzato il vettore giornaliero
# dell'ozono)
o3AVEm<-matrix(finale1$o3.AVE,1,911)
o3fdcost<- fd(o3AVEm, create.constant.basis(dayrange))

#  basis di tipo “fourier” con numero di basis pari a 5:
```

```
smallnbasis <- 5
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)
o3fd5<- data2fd(o3m, daytime, smallbasis)

#  basis di tipo “fourier” con numero di basis pari a 7:
smallnbasis <- 7
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)
o3fd7<- data2fd(o3m, daytime, smallbasis)

#  basis di tipo “fourier” con numero di basis pari a 9:
smallnbasis <- 9
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)
o3fd9<- data2fd(o3m, daytime, smallbasis)

#  basis di tipo “fourier” con numero di basis pari a 23:
# (con 23 la basis si dice “satura” perchè è l'estremo superiore dell'intervallo (0,23)  )
smallnbasis <- 23
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)
o3fd23<- data2fd(o3m, daytime, smallbasis)
```

Regressione con basis costante:

```
#  scelta dei tempi di osservazione :
daytime   <- (1:24)
dayrange  <- c(0,24)
dayperiod <- 24

#
smallbasismat <- eval.basis(daytime, smallbasis)

#  covariate:

# ossigeno
o3AVEm<-matrix(finale1$o3.AVE,1,911)
o3fdcost<- fd(o3AVEm, create.constant.basis(dayrange))

# costante
```

```
constantfd <- fd(matrix(1,1,911), create.constant.basis(dayrange))

# temperatura media giornaliera
tempm<-matrix(finale1$temmean,1,911)
tempfd <- fd(tempm, create.constant.basis(dayrange))

# indicatore progressivo giornaliero
tempom<-matrix(finale1$tempo,1,911)
tempofd <- fd(tempom, create.constant.basis(dayrange))

# pm10
pm10m<-matrix(finale1$pm10.AVE,1,911)
pm10fd <- fd(pm10m, create.constant.basis(dayrange))

# indicatore del giorno di festa
festam<-matrix(finale1$festa01,1,911)
festafd <- fd(festam, create.constant.basis(dayrange))

# variazione di temperatura
Tscartom<-matrix(finale1$Tscarto,1,911)
Tscartofd <- fd(Tscartom, create.constant.basis(dayrange))

# indicatore del giorno della settimana
wdaym<-matrix(finale1$wday,1,911)
wdayfd <- fd(wdaym, create.constant.basis(dayrange))

# numero covariate:
p<-8

# creazione vettore di tipo “lista” delle covariate funzionali:
xfdlist <- vector("list",p)
xfdlist[[1]] <- constantfd
xfdlist[[2]] <- o3fdcost
xfdlist[[3]] <- tempfd
xfdlist[[4]] <- tempofd
xfdlist[[5]] <- pm10fd
xfdlist[[6]] <- festafd
xfdlist[[7]] <- Tscartofd
```

```
xfdlist[[8]] <- wdayfd

# creazione oggetto “parametri funzionali”, contenente i coefficienti della regressione funzionale:
betalist   <- vector("list",p)
#  il primo regressore è la costante:
betabasis1 <- create.constant.basis(dayrange)
betafd1    <- fd(0, betabasis1)
betafdPar1 <- fdPar(betafd1)
betalist[[1]] <- betafdPar1
#  il secondo regressore
betabasis2 <- create.constant.basis(dayrange)
betafd2    <- fd(0, betabasis1)
betafdPar2 <- fdPar(betafd2)
betalist[[2]] <- betafdPar2
#
betabasis3 <- create.constant.basis(dayrange)
betafd3    <- fd(0, betabasis3)
betafdPar3 <- fdPar(betafd3)
betalist[[3]] <- betafdPar3
#
betabasis4 <- create.constant.basis(dayrange)
betafd4    <- fd(0, betabasis4)
betafdPar4 <- fdPar(betafd4)
betalist[[4]] <- betafdPar4
#
betabasis5 <- create.constant.basis(dayrange)
betafd5    <- fd(0, betabasis5)
betafdPar5 <- fdPar(betafd5)
betalist[[5]] <- betafdPar5
#
betabasis6 <- create.constant.basis(dayrange)
betafd6    <- fd(0, betabasis6)
betafdPar6 <- fdPar(betafd6)
betalist[[6]] <- betafdPar6
#
betabasis7 <- create.constant.basis(dayrange)
betafd7    <- fd(0, betabasis7)
betafdPar7 <- fdPar(betafd7)
```

```
betalist[[7]] <- betafdPar7
#
betabasis8 <- create.constant.basis(dayrange)
betafd8    <- fd(0, betabasis8)
betafdPar8 <- fdPar(betafd8)
betalist[[8]] <- betafdPar8


#  regressione funzionale:
fRegressList <- fRegress(pzlog, xfdlist, betalist)

# coefficienti:
fRegressList$betaestlist[[1]]$fd$coefs
fRegressList$betaestlist[[2]]$fd$coefs
fRegressList$betaestlist[[3]]$fd$coefs
fRegressList$betaestlist[[4]]$fd$coefs
fRegressList$betaestlist[[5]]$fd$coefs
fRegressList$betaestlist[[6]]$fd$coefs
fRegressList$betaestlist[[7]]$fd$coefs
fRegressList$betaestlist[[8]]$fd$coefs

betaestlist<-fRegressList$betaestlist

# valori stimati dal modello:
pzloghat<-fRegressList$yhatfdobj

#  calcolo del R^2 :
covmat <- var(cbind(pzlog, pzloghat))
Rsqrd<-covmat[1,2]^2/(covmat[1,1]*covmat[2,2])
Rsqrd

#  calcolo della matrice SigmaE:
resid  <- pzlog - pzloghat
SigmaE <- mean(resid^2)
SigmaE <- SigmaE*diag(rep(1,911))

#  calcolo degli standard error:
stderrList <- fRegress.stderr(fRegressList, NULL, SigmaE)
```

```
betastderrlist <- stderrList$betastderrlist

#  standard error dei coefficienti:
betastderrlist[[1]]$coefs
betastderrlist[[2]]$coefs
betastderrlist[[3]]$coefs
betastderrlist[[4]]$coefs
betastderrlist[[5]]$coefs
betastderrlist[[6]]$coefs
betastderrlist[[7]]$coefs
betastderrlist[[8]]$coefs

# calcolo dei test t per la significatività dei coefficienti:
t1<-fRegressList$betaestlist[[1]]$fd$coefs/betastderrlist[[1]]$coefs
t2<-fRegressList$betaestlist[[2]]$fd$coefs/betastderrlist[[2]]$coefs
t3<-fRegressList$betaestlist[[3]]$fd$coefs/betastderrlist[[3]]$coefs
t4<-fRegressList$betaestlist[[4]]$fd$coefs/betastderrlist[[4]]$coefs
t5<-fRegressList$betaestlist[[5]]$fd$coefs/betastderrlist[[5]]$coefs
t6<-fRegressList$betaestlist[[6]]$fd$coefs/betastderrlist[[6]]$coefs
t7<-fRegressList$betaestlist[[7]]$fd$coefs/betastderrlist[[7]]$coefs
t8<-fRegressList$betaestlist[[8]]$fd$coefs/betastderrlist[[8]]$coefs

# stampa dei valori dei test t:
list(t1,t2,t3,t4,t5,t6,t7,t8)
```

Regressione con basis di tipo “fourier”:

NOTA: viene presentata solo la regressione con numero di basis pari a 9, in quanto le altre possono essere facilmente ottenute inserendo nel parametro “nbasis” il numero di basis voluto.

```
#  scelta dei tempi di osservazione:
daytime   <- (1:24)
dayrange  <- c(0,24)
dayperiod <- 24

#  scelta della basis “fourier”:
nbasis     <- 9
```

```
daybasis <- create.fourier.basis(dayrange, nbasis, dayperiod)

harmaccelLfd <- vec2Lfd(c(0,(2*pi/24)^2,0), dayrange)

smallnbasis <- nbasis
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)

smallbasismat <- eval.basis(daytime, smallbasis)

# covariate:
# (solo l'ozono ha basis “fourier”, le altre covariate mantengono la basis costante)

o3fd<- data2fd(o3m, daytime, smallbasis)

constantfd <- fd(matrix(1,1,911), create.constant.basis(dayrange))

tempm<-matrix(finale1$temmean,1,911)
tempfd <- fd(tempm, create.constant.basis(dayrange))

tempom<-matrix(finale1$tempo,1,911)
tempofd <- fd(tempom, create.constant.basis(dayrange))

pm10m<-matrix(finale1$pm10.AVE,1,911)
pm10fd <- fd(pm10m, create.constant.basis(dayrange))

festam<-matrix(finale1$festa01,1,911)
festafd <- fd(festam, create.constant.basis(dayrange))

Tscartom<-matrix(finale1$Tscarto,1,911)
Tscartofd <- fd(Tscartom, create.constant.basis(dayrange))

wdaym<-matrix(finale1$wday,1,911)
wdayfd <- fd(wdaym, create.constant.basis(dayrange))

# numero covariate:
p<-8

# creazione vettore di tipo “lista” delle covariate funzionali:
```

```
xfdlist <- vector("list",p)
xfdlist[[1]] <- constantfd
xfdlist[[2]] <- o3fd
xfdlist[[3]] <- tempfd
xfdlist[[4]] <- tempofd
xfdlist[[5]] <- pm10fd
xfdlist[[6]] <- festafd
xfdlist[[7]] <- Tscartofd
xfdlist[[8]] <- wdayfd

# creazione oggetto “parametri funzionali”, contenente i coefficienti della regressione funzionale:
betalist   <- vector("list",p)
#  il primo regressore è la costante:
betabasis1 <- create.constant.basis(dayrange)
betafd1    <- fd(0, betabasis1)
betafdPar1 <- fdPar(betafd1)
betalist[[1]] <- betafdPar1
#  per coefficiente del regressore ozono viene utilizzata la basis “fourier”:
nbetabasis  <- nbasis
betabasis2  <- create.fourier.basis(dayrange, nbetabasis)
betafd2     <- fd(matrix(0,nbetabasis,1), betabasis2)
lambda      <- 10
betafdPar2  <- fdPar(betafd2, harmaccelLfd, lambda)
betalist[[2]] <- betafdPar2
#
betabasis3 <- create.constant.basis(dayrange)
betafd3    <- fd(0, betabasis3)
betafdPar3 <- fdPar(betafd3)
betalist[[3]] <- betafdPar3
#
betabasis4 <- create.constant.basis(dayrange)
betafd4    <- fd(0, betabasis4)
betafdPar4 <- fdPar(betafd4)
betalist[[4]] <- betafdPar4
#
betabasis5 <- create.constant.basis(dayrange)
betafd5    <- fd(0, betabasis5)
betafdPar5 <- fdPar(betafd5)
```

```
betalist[[5]] <- betafdPar5
#
betabasis6 <- create.constant.basis(dayrange)
betafd6    <- fd(0, betabasis6)
betafdPar6 <- fdPar(betafd6)
betalist[[6]] <- betafdPar6
#
betabasis7 <- create.constant.basis(dayrange)
betafd7    <- fd(0, betabasis7)
betafdPar7 <- fdPar(betafd7)
betalist[[7]] <- betafdPar7
#
betabasis8 <- create.constant.basis(dayrange)
betafd8    <- fd(0, betabasis8)
betafdPar8 <- fdPar(betafd8)
betalist[[8]] <- betafdPar8


# regressione funzionale:
fRegressList <- fRegress(pzlog, xfdlist, betalist)

# coefficienti:
fRegressList$betaestlist[[1]]$fd$coefs
fRegressList$betaestlist[[2]]$fd$coefs
fRegressList$betaestlist[[3]]$fd$coefs
fRegressList$betaestlist[[4]]$fd$coefs
fRegressList$betaestlist[[5]]$fd$coefs
fRegressList$betaestlist[[6]]$fd$coefs
fRegressList$betaestlist[[7]]$fd$coefs
fRegressList$betaestlist[[8]]$fd$coefs

betaestlist<-fRegressList$betaestlist

# valori stimati dal modello:
pzloghat<-fRegressList$yhatfdobj

#  calcolo del R^2 :
covmat <- var(cbind(pzlog, pzloghat))
```

```
Rsqrd<-covmat[1,2]^2/(covmat[1,1]*covmat[2,2])
Rsqrd

# calcolo della matrice SigmaE:
resid  <- pzlog - pzloghat
SigmaE <- mean(resid^2)
SigmaE <- SigmaE*diag(rep(1,911))

#  calcolo degli standard error:
stderrList <- fRegress.stderr(fRegressList, NULL, SigmaE)
betastderrlist <- stderrList$betastderrlist

# standard error:
betastderrlist[[1]]$coefs
betastderrlist[[2]]$coefs
betastderrlist[[3]]$coefs
betastderrlist[[4]]$coefs
betastderrlist[[5]]$coefs
betastderrlist[[6]]$coefs
betastderrlist[[7]]$coefs
betastderrlist[[8]]$coefs

# calcolo dei test t per la significatività dei coefficienti:
t1<-fRegressList$betaestlist[[1]]$fd$coefs/betastderrlist[[1]]$coefs
t2<-fRegressList$betaestlist[[2]]$fd$coefs/betastderrlist[[2]]$coefs
t3<-fRegressList$betaestlist[[3]]$fd$coefs/betastderrlist[[3]]$coefs
t4<-fRegressList$betaestlist[[4]]$fd$coefs/betastderrlist[[4]]$coefs
t5<-fRegressList$betaestlist[[5]]$fd$coefs/betastderrlist[[5]]$coefs
t6<-fRegressList$betaestlist[[6]]$fd$coefs/betastderrlist[[6]]$coefs
t7<-fRegressList$betaestlist[[7]]$fd$coefs/betastderrlist[[7]]$coefs
t8<-fRegressList$betaestlist[[8]]$fd$coefs/betastderrlist[[8]]$coefs

# stampa dei valori dei test t:
list(t1,t2,t3,t4,t5,t6,t7,t8)
```

Regressione con basis di tipo “bspline”:

```
# scelta dei tempi di osservazione:
daytime<-c(0:23)
dayrange<-c(0,23)
dayrange

# creazione basis “bspline” con numero basis pari a 9:
nbasis<-9
daybasis<-create.bspline.basis(dayrange,nbasis)

# covariate:
# (solo l'ozono ha basis “bspline”, le altre covariate mantengono la basis costante)

o3fdbspline<-data2fd(o3m,daytime,daybasis)

constantfd <- fd(matrix(1,1,911), create.constant.basis(dayrange))

tempm<-matrix(finale1$temmean,1,911)
tempfd <- fd(tempm, create.constant.basis(dayrange))

tempom<-matrix(finale1$tempo,1,911)
tempofd <- fd(tempom, create.constant.basis(dayrange))

pm10m<-matrix(finale1$pm10.AVE,1,911)
pm10fd <- fd(pm10m, create.constant.basis(dayrange))

festam<-matrix(finale1$festa01,1,911)
festafd <- fd(festam, create.constant.basis(dayrange))

Tscartom<-matrix(finale1$Tscarto,1,911)
Tscartofd <- fd(Tscartom, create.constant.basis(dayrange))

wdaym<-matrix(finale1$wday,1,911)
wdayfd <- fd(wdaym, create.constant.basis(dayrange))

# numero covariate:
p<-8

# creazione vettore di tipo “lista” delle covariate funzionali:
```

```
xfdlist <- vector("list",p)
xfdlist[[1]] <- constantfd
xfdlist[[2]] <- o3fdbspline
xfdlist[[3]] <- tempfd
xfdlist[[4]] <- tempofd
xfdlist[[5]] <- pm10fd
xfdlist[[6]] <- festafd
xfdlist[[7]] <- Tscartofd
xfdlist[[8]] <- wdayfd

# creazione oggetto “parametri funzionali”, contenente i coefficienti della regressione funzionale:
betalist   <- vector("list",p)
#  il primo regressore è la costante:
betabasis1 <- create.constant.basis(dayrange)
betafd1    <- fd(0, betabasis1)
betafdPar1 <- fdPar(betafd1)
betalist[[1]] <- betafdPar1
#
nbetabasis  <- nbasis
betabasis2  <- create.bspline.basis(dayrange,nbasis)
betafd2     <- fd(matrix(0,nbetabasis,1), betabasis2)
betafdPar2  <- fdPar(betafd2)
betalist[[2]] <- betafdPar2
#
betabasis3 <- create.constant.basis(dayrange)
betafd3    <- fd(0, betabasis3)
betafdPar3 <- fdPar(betafd3)
betalist[[3]] <- betafdPar3
#
betabasis4 <- create.constant.basis(dayrange)
betafd4    <- fd(0, betabasis4)
betafdPar4 <- fdPar(betafd4)
betalist[[4]] <- betafdPar4
#
betabasis5 <- create.constant.basis(dayrange)
betafd5    <- fd(0, betabasis5)
betafdPar5 <- fdPar(betafd5)
betalist[[5]] <- betafdPar5
```

```
#
betabasis6 <- create.constant.basis(dayrange)
betafd6    <- fd(0, betabasis6)
betafdPar6 <- fdPar(betafd6)
betalist[[6]] <- betafdPar6
#
betabasis7 <- create.constant.basis(dayrange)
betafd7    <- fd(0, betabasis7)
betafdPar7 <- fdPar(betafd7)
betalist[[7]] <- betafdPar7
#
betabasis8 <- create.constant.basis(dayrange)
betafd8    <- fd(0, betabasis8)
betafdPar8 <- fdPar(betafd8)
betalist[[8]] <- betafdPar8


#  regressione funzionale:
fRegressList <- fRegress(pzlog, xfdlist, betalist)

# coefficienti:
fRegressList$betaestlist[[1]]$fd$coefs
fRegressList$betaestlist[[2]]$fd$coefs
fRegressList$betaestlist[[3]]$fd$coefs
fRegressList$betaestlist[[4]]$fd$coefs
fRegressList$betaestlist[[5]]$fd$coefs
fRegressList$betaestlist[[6]]$fd$coefs
fRegressList$betaestlist[[7]]$fd$coefs
fRegressList$betaestlist[[8]]$fd$coefs

betaestlist<-fRegressList$betaestlist

# valori stimati dal modello:
pzloghat<-fRegressList$yhatfdobj

#  calcolo del R^2 :
covmat <- var(cbind(pzlog, pzloghat))
Rsqrd<-covmat[1,2]^2/(covmat[1,1]*covmat[2,2])
```

```
Rsqrd

#  calcolo della matrice SigmaE:
resid  <- pzlog - pzloghat
SigmaE <- mean(resid^2)
SigmaE <- SigmaE*diag(rep(1,911))

#  calcolo degli standard error:
stderrList <- fRegress.stderr(fRegressList, NULL, SigmaE)

betastderrlist <- stderrList$betastderrlist

#  standard error dei coefficienti:
betastderrlist[[1]]$coefs
betastderrlist[[2]]$coefs
betastderrlist[[3]]$coefs
betastderrlist[[4]]$coefs
betastderrlist[[5]]$coefs
betastderrlist[[6]]$coefs
betastderrlist[[7]]$coefs
betastderrlist[[8]]$coefs

# calcolo dei test t per la significatività dei coefficienti:
t1<-fRegressList$betaestlist[[1]]$fd$coefs/betastderrlist[[1]]$coefs
t2<-fRegressList$betaestlist[[2]]$fd$coefs/betastderrlist[[2]]$coefs
t3<-fRegressList$betaestlist[[3]]$fd$coefs/betastderrlist[[3]]$coefs
t4<-fRegressList$betaestlist[[4]]$fd$coefs/betastderrlist[[4]]$coefs
t5<-fRegressList$betaestlist[[5]]$fd$coefs/betastderrlist[[5]]$coefs
t6<-fRegressList$betaestlist[[6]]$fd$coefs/betastderrlist[[6]]$coefs
t7<-fRegressList$betaestlist[[7]]$fd$coefs/betastderrlist[[7]]$coefs
t8<-fRegressList$betaestlist[[8]]$fd$coefs/betastderrlist[[8]]$coefs

# stampa dei valori dei test t:
list(t1,t2,t3,t4,t5,t6,t7,t8)
```

Regressione lineare:

summary(lm(pzlog~finale1$o3.AVE+finale1$temmean+finale1$tempo+finale1$pm10.AVE+finale1$festa01+finale1$Tscarto+finale1$wday))

APPENDICE B: funzione regressione.

La funzione qui descritta calcola velocemente una regressione funzionale ed i suoi risultati utilizzati in questa tesi. Nello specifico, è riferita alla regressione con variabile risposta log(ricoveri) e con gli 8 regressori già utilizzati in precedenza, con una base “Fourier” per l'ozono.
Lo scopo è rendere meno onerosa per il programmatore la scrittura del codice per calcolare regressioni con diverso numero di funzioni di base e confrontarne i risultati.
Ad esempio, per una regressione con numero di funzioni di base pari a 23 è sufficiente scrivere il comando:
regressione(23)
e la funzione produce un oggetto dal nome “risultati” (di tipo “lista”) contenente il numero di funzini di base scelto all'inizio, il valore del R^2, i coefficienti dei regressori, gli standard error dei regressori ed il valore dei test t sui coefficienti.

```
### dati preliminari:
# scelta dei tempi di osservazione:
daytime   <- (1:24)
dayrange  <- c(0,24)
dayperiod <- 24
#  scelta del numero di basis:
nbasis     <- 9

### funzione :

regressione<-function(nbasis){

# creazione della basis “fourier”:
daybasis <- create.fourier.basis(dayrange, nbasis, dayperiod)

harmaccelLfd <- vec2Lfd(c(0,(2*pi/24)^2,0), dayrange)

smallnbasis <- nbasis
smallbasis  <- create.fourier.basis(dayrange, smallnbasis)


smallbasismat <- eval.basis(daytime, smallbasis)
```

```
# covariate:
# (solo l'ozono ha basis “fourier”, le altre covariate mantengono la basis costante)

o3fd<- data2fd(o3m, daytime, smallbasis)

constantfd <- fd(matrix(1,1,911), create.constant.basis(dayrange))

tempm<-matrix(finale1$temmean,1,911)
tempfd <- fd(tempm, create.constant.basis(dayrange))

tempom<-matrix(finale1$tempo,1,911)
tempofd <- fd(tempom, create.constant.basis(dayrange))

pm10m<-matrix(finale1$pm10.AVE,1,911)
pm10fd <- fd(pm10m, create.constant.basis(dayrange))

festam<-matrix(finale1$festa01,1,911)
festafd <- fd(festam, create.constant.basis(dayrange))

Tscartom<-matrix(finale1$Tscarto,1,911)
Tscartofd <- fd(Tscartom, create.constant.basis(dayrange))

wdaym<-matrix(finale1$wday,1,911)
wdayfd <- fd(wdaym, create.constant.basis(dayrange))

# numero covariate:
p<-8

# creazione vettore di tipo “lista” delle covariate funzionali:
xfdlist <- vector("list",p)
xfdlist[[1]] <- constantfd
xfdlist[[2]] <- o3fd
xfdlist[[3]] <- tempfd
xfdlist[[4]] <- tempofd
xfdlist[[5]] <- pm10fd
xfdlist[[6]] <- festafd
```

```
xfdlist[[7]] <- Tscartofd
xfdlist[[8]] <- wdayfd

# creazione oggetto “parametri funzionali”, contenente i coefficienti della regressione funzionale:
betalist   <- vector("list",p)
#  il primo regressore è la costante:
betabasis1 <- create.constant.basis(dayrange)
betafd1    <- fd(0, betabasis1)
betafdPar1 <- fdPar(betafd1)
betalist[[1]] <- betafdPar1
#  per il coefficiente del regressore ozono viene utilizzata la basis Fourier”:
nbetabasis  <- nbasis
betabasis2  <- create.fourier.basis(dayrange, nbetabasis)
betafd2     <- fd(matrix(0,nbetabasis,1), betabasis2)
lambda      <- 10
betafdPar2  <- fdPar(betafd2, harmaccelLfd, lambda)
betalist[[2]] <- betafdPar2
#
betabasis3 <- create.constant.basis(dayrange)
betafd3    <- fd(0, betabasis3)
betafdPar3 <- fdPar(betafd3)
betalist[[3]] <- betafdPar3
#
betabasis4 <- create.constant.basis(dayrange)
betafd4    <- fd(0, betabasis4)
betafdPar4 <- fdPar(betafd4)
betalist[[4]] <- betafdPar4
#
betabasis5 <- create.constant.basis(dayrange)
betafd5    <- fd(0, betabasis5)
betafdPar5 <- fdPar(betafd5)
betalist[[5]] <- betafdPar5
#
betabasis6 <- create.constant.basis(dayrange)
betafd6    <- fd(0, betabasis6)
betafdPar6 <- fdPar(betafd6)
betalist[[6]] <- betafdPar6
#
```

```
betabasis7 <- create.constant.basis(dayrange)
betafd7    <- fd(0, betabasis7)
betafdPar7 <- fdPar(betafd7)
betalist[[7]] <- betafdPar7
#
betabasis8 <- create.constant.basis(dayrange)
betafd8    <- fd(0, betabasis8)
betafdPar8 <- fdPar(betafd8)
betalist[[8]] <- betafdPar8



# regressione funzionale:
fRegressList <- fRegress(pzlog, xfdlist, betalist)

# coefficienti:
coef1<-fRegressList$betaestlist[[1]]$fd$coefs
coef2<-fRegressList$betaestlist[[2]]$fd$coefs
coef3<-fRegressList$betaestlist[[3]]$fd$coefs
coef4<-fRegressList$betaestlist[[4]]$fd$coefs
coef5<-fRegressList$betaestlist[[5]]$fd$coefs
coef6<-fRegressList$betaestlist[[6]]$fd$coefs
coef7<-fRegressList$betaestlist[[7]]$fd$coefs
coef8<-fRegressList$betaestlist[[8]]$fd$coefs

coef<-list(coef1,coef2,coef3,coef4,coef5,coef6,coef7,coef8)

betaestlist<-fRegressList$betaestlist

# valori stimati dal modello:
pzloghat<-fRegressList$yhatfdobj

#  calcolo del R^2 :
covmat <- var(cbind(pzlog, pzloghat))
Rsqrd<-covmat[1,2]^2/(covmat[1,1]*covmat[2,2])

# calcolo della matrice SigmaE:
resid  <- pzlog - pzloghat
SigmaE <- mean(resid^2)
```

```
SigmaE <- SigmaE*diag(rep(1,911))

#  calcolo degli standard error:
stderrList <- fRegress.stderr(fRegressList, NULL, SigmaE)
betastderrlist <- stderrList$betastderrlist

# standard error:
sd1<-betastderrlist[[1]]$coefs
sd2<-betastderrlist[[2]]$coefs
sd3<-betastderrlist[[3]]$coefs
sd4<-betastderrlist[[4]]$coefs
sd5<-betastderrlist[[5]]$coefs
sd6<-betastderrlist[[6]]$coefs
sd7<-betastderrlist[[7]]$coefs
sd8<-betastderrlist[[8]]$coefs

sd<-list(sd1,sd2,sd3,sd4,sd5,sd6,sd7,sd8)

# calcolo dei test t per la significatività dei coefficienti:
t1<-fRegressList$betaestlist[[1]]$fd$coefs/betastderrlist[[1]]$coefs
t2<-fRegressList$betaestlist[[2]]$fd$coefs/betastderrlist[[2]]$coefs
t3<-fRegressList$betaestlist[[3]]$fd$coefs/betastderrlist[[3]]$coefs
t4<-fRegressList$betaestlist[[4]]$fd$coefs/betastderrlist[[4]]$coefs
t5<-fRegressList$betaestlist[[5]]$fd$coefs/betastderrlist[[5]]$coefs
t6<-fRegressList$betaestlist[[6]]$fd$coefs/betastderrlist[[6]]$coefs
t7<-fRegressList$betaestlist[[7]]$fd$coefs/betastderrlist[[7]]$coefs
t8<-fRegressList$betaestlist[[8]]$fd$coefs/betastderrlist[[8]]$coefs

t<-list(t1,t2,t3,t4,t5,t6,t7,t8)

D1<-t

# creazione di un oggetto “list” contenente i risultati della regressione:
risultati<-list(nbasis,Rsqrd,coef,sd,t)



return(risultati)
# risultati[1] restituisce il numero di basis scelto all'inizio,
```

risultati[2] restituisce il valore del R^2 ,
risultati[3] restituisce i coefficienti dei regressori,
risultati[4] restituisce gli standard error dei regressori,
risultati[5] restituisce il valore dei test t sui coefficienti.

}