

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per l'Economia e l'Impresa



RELAZIONE FINALE

**COLTIVAZIONI AGRICOLE E IMPIANTI FOTOVOLTAICI.
ANALISI STATISTICA DEI DIFFERENZIALI DI CRESCITA IN
CONDIZIONI DI OMBREGGIAMENTO VARIABILE**

Relatore Prof. Luigi Grossi
Dipartimento di Scienze Statistiche

Laureanda: Benedetta Piacenza
Matricola N 1232345

Anno Accademico 2021/2022

ABSTRACT

Il presente lavoro di tesi ha come obiettivo l'analisi statistica dei differenziali di crescita per piante coltivate in condizioni di ombreggiamento variabile in impianti agrivoltaici. Questo è stato realizzato tramite l'applicazione di tecniche di regressione, quali la regressione lineare multipla con trend polinomiale e la regressione quantilica, realizzate tramite il software statistico R Studio.

Il tema principale in questo elaborato è l'agrivoltaico, cioè un sistema combinato di produzione agricola e di generazione di energia tramite impianti fotovoltaici, realizzate sul medesimo terreno. Nel corso della presente tesi verranno esposti i principi fondamentali di questo metodo di produzione di energia, descrivendo come è strutturato un tipico impianto, citando brevemente la storia della tecnologia e i principali esperimenti effettuati in questo ambito riferendosi a un'analisi della letteratura. Successivamente sono stati applicati a questo ambito alcuni strumenti statistici, come i modelli di regressione, per valutare le possibili differenze di crescita in diverse condizioni di ombreggiamento.

Attraverso questa analisi si è giunti alla conclusione che non vi sono differenze significative nella crescita delle piante per le due condizioni di illuminazione, piena luce e in ombra. Quindi, tenendo conto dei molti benefici indotti da un impianto agrivoltaico, che verranno descritti in seguito, si conclude che un sistema di questo tipo è preferibile ad un impianto fotovoltaico convenzionale.

INDICE

INTRODUZIONE	1
CAPITOLO 1	3
L'AGRIVOLTAICO	3
1.1 Premessa.....	3
1.2 Definizione	4
1.3 Storia.....	5
1.4 Descrizione dell'impianto.....	6
1.5 Analisi della letteratura.....	9
CAPITOLO 2	13
STRUMENTI STATISTICI	13
2.1 Modelli di regressione per serie storiche.....	13
2.1.1 <i>Le serie storiche</i>	13
2.1.2 <i>Approccio classico vs moderno</i>	14
2.1.3 <i>Le componenti di una serie storica</i>	14
2.1.4 <i>Modelli di regressione lineare</i>	15
2.1.6 <i>Stima del trend mediante funzioni polinomiali</i>	17
2.1.7 <i>Il confronto tra R^2 per la scelta del grado del polinomio</i>	18
2.2 Regressione quantilica.....	18
2.2.1 <i>Il modello QR più semplice: Il caso del regressore dummy</i>	20
2.2.2 <i>Modello QR con regressore nominale</i>	21
2.2.3 <i>Modello QR con regressore quantitativo</i>	21
2.3 ANOVA e ANCOVA.....	21
2.3.1 <i>ANOVA a un fattore</i>	22
2.3.2 <i>ANOVA a due fattori</i>	23
2.3.3 <i>ANCOVA</i>	25
CAPITOLO 3	27
APPLICAZIONE AI DATI	27
3.1 Coltura erba medica.....	27
3.1.1 <i>Descrizione dei dati</i>	27
3.1.2 <i>Modelli di regressione con trend polinomiale</i>	30
3.1.3 <i>Regressione quantilica</i>	33
3.2 Coltura fagioli.....	36

3.2.1	<i>Descrizione dei dati</i>	36
3.2.2	<i>Modelli di regressione con trend polinomiale</i>	38
3.2.3	<i>Regressione quantilica</i>	44
3.3	Coltura patate	51
3.3.1	<i>Descrizione dei dati</i>	51
3.3.2	<i>Modelli di regressione con trend polinomiale</i>	53
3.3.3	<i>Regressione quantilica</i>	59
CONCLUSIONI		65
BIBLIOGRAFIA		67
APPENDICE		69
	Grafici erba medica	69
	Grafici fagioli	71
	Grafici patate	73

INTRODUZIONE

L'argomento affrontato in questo elaborato è il tema dell'agrivoltaico, ossia un sistema combinato di produzione agricola e di generazione di energia tramite impianti fotovoltaici, realizzate sul medesimo terreno. Il principio di funzionamento di un impianto di questo tipo consiste nel posizionare su terreni agricoli dei pannelli fotovoltaici di un'altezza tale da rendere possibile la coltivazione di piante al di sotto di essi. I benefici che questo sistema può fornire sono: riduzione dell'uso del suolo rispetto ai tradizionali impianti, eliminazione della concorrenza con la produzione a fini alimentari, riduzione dell'evapotraspirazione dovuta all'ombra creata dai pannelli con conseguente diminuzione del consumo di acqua, infine essendo una fonte di energia rinnovabile non produce emissioni di gas serra come le tradizionali fonti fossili.

L'obiettivo della presente tesi è l'analisi statistica dell'andamento temporale di alcuni parametri correlati alla crescita delle piante, in diverse condizioni di luce o ombra, tramite due tipologie di modelli.

L'elaborato è strutturato in 3 capitoli: nel primo capitolo, viene presentato e descritto in dettaglio il concetto di agrivoltaico, il funzionamento di un impianto e un'analisi della letteratura riguardo i benefici elencati sopra e gli esperimenti condotti fino ad oggi sull'argomento.

Nel secondo capitolo vengono presentati i modelli statistici utilizzati per l'analisi: il modello di regressione per serie storiche (con trend cubico), per spiegare la relazione in media tra una variabile dipendente detta risposta e un insieme di variabili dipendenti dette regressori, la regressione quantilica, che ha la stessa funzione ma per diversi quantili della distribuzione della variabile risposta e l'analisi della varianza (ANOVA) a uno e due fattori e l'analisi della covarianza (ANCOVA), due tecniche utilizzate per confrontare due o più gruppi e la variabilità interna e tra essi.

Infine, nel terzo capitolo vengono applicati i primi due modelli descritti nel capitolo precedente a tre dataset al fine di valutare le differenze nella crescita di alcuni parametri nelle diverse condizioni in cui sono state coltivate le piante. I dati sono relativi alle seguenti tipologie di piante: erba medica, fagioli e patate, per ciascuna di esse verranno analizzati diversi parametri di crescita che verranno elencati nel relativo paragrafo.

In conclusione, i risultati ottenuti evidenziano che non vi sono differenze significative nello sviluppo dei parametri per ciascuna tipologia di pianta., ciò implica che la crescita nelle due condizioni di illuminazione, piena luce e in ombra, sono simili. Pertanto, visti i molteplici benefici indotti da un impianto agrivoltaico, si conclude che un sistema di questo tipo è preferibile ad un impianto fotovoltaico convenzionale.

CAPITOLO 1

L'AGRIVOLTAICO

1.1 Premessa

La domanda di energia mondiale è in continua crescita a causa dell'aumento della popolazione e dell'uso sempre maggiore da parte dell'industria. Allo stesso tempo la minaccia del riscaldamento globale sta cambiando le modalità di produzione dell'energia verso fonti rinnovabili. I governi di tutto il mondo ripongono grande fiducia nelle energie rinnovabili come importanti tecnologie per ridurre i problemi ambientali legati all'energia, in particolare alle emissioni di CO₂. Nel 2015 tutti gli stati membri delle nazioni unite (ONU) hanno adottato l'Agenda 2030 per lo Sviluppo Sostenibile [7], la quale definisce 17 obiettivi di sviluppo sostenibile (*Sustainable Development Goals*). Nell'obiettivo 7 dell'Agenda 2030, “Assicurare a tutti l'accesso a sistemi di energia economici, affidabili, sostenibili e moderni”, si trovano, ad esempio, indicati i seguenti traguardi:

7.2 Aumentare considerevolmente entro il 2030 la quota di energie rinnovabili nel consumo totale di energia

7.3 Raddoppiare entro il 2030 il tasso globale di miglioramento dell'efficienza energetica

Le energie rinnovabili sono quelle forme di energia generate da fonti che, per loro caratteristica intrinseca, si rigenerano o non sono “esauribili” nella scala dei tempi “umani” e il cui utilizzo non pregiudica le risorse naturali per le generazioni future, in quanto esse rinnovano la loro disponibilità in tempi estremamente brevi. Sono dunque fonti di energia alternative alle tradizionali fonti fossili e inoltre possiedono la caratteristica di avere un basso impatto ambientale, poiché non immettono nell'atmosfera sostanze nocive per l'uomo e/o per l'ambiente.

Secondo l'ISES (International Solar Energy Society) le fonti energetiche rinnovabili sono classificate come segue:

- Energia eolica
- Energia solare: termico e fotovoltaico
- Energia idraulica: impianti elettrici di piccola taglia (small hydro)
- Biomasse: biomassa solida, prodotti animali, gas/liquidi da biomassa, legna, biodisel
- Energia geotermica
- Energia delle onde e delle maree

Tra tutte le energie pulite e rinnovabili, l'energia solare è la fonte più disponibile e abbondante. Il fotovoltaico, in particolare, rappresenta quella che si sta sviluppando più rapidamente e si prevede che continuerà a crescere in modo esponenziale nei prossimi anni, svolgendo un ruolo sempre più importante per il raggiungimento degli obiettivi di energia rinnovabile stabiliti dai governi nazionali.

L'energia solare fotovoltaica è una forma di energia rinnovabile molto interessante per le seguenti ragioni [12]:

- Significative emissioni di carbonio sono evitate
- I pannelli solari hanno un tempo di vita ampio (20-30 anni)
- È una fonte di energia stabile, abbondante e a basso costo
- I pannelli sono efficienti nel catturare l'energia solare rispetto alla fotosintesi

Tuttavia, gli impianti fotovoltaici presentano anche degli elementi a loro sfavore. Infatti, la produzione di energia elettrica rinnovabile tramite i classici impianti fotovoltaici installati a terra presenta lo svantaggio di avere bisogno di ampie superfici per l'installazione dei pannelli, sottraendo così terreni alla produzione agricola e generando conseguentemente un conflitto tra la produzione di energia e la produzione alimentare. Questo risulta particolarmente problematico nelle regioni con spazi limitati di terra o ad alti tassi di popolazione. Una soluzione al suddetto problema è fornita dall'agrivoltaico.

1.2 Definizione

Il termine agrivoltaico indica un sistema che combina, sulla stessa area di terra, l'attività agricola con la produzione di energia elettrica tramite pannelli fotovoltaici installati a diversi metri dal suolo [1]. L'obiettivo primario del sistema agrivoltaico consiste nella preservazione dei terreni all'utilizzo agricolo. Infatti, negli impianti agrivoltaici i pannelli fotovoltaici sono montati ad un'altezza da terra sufficiente a consentire pratiche di coltivazione convenzionali sul terreno sottostante. Inoltre, l'ombreggiamento causato dai moduli fotovoltaici aiuta a diminuire l'evapotraspirazione, riducendo così il consumo di acqua, e ciò si rivela vantaggioso soprattutto nella stagione secca.

La presenza dei pannelli fotovoltaici, in analogia agli alberi nei sistemi agroforestali, protegge le colture dal calore eccessivo e mitiga la temperatura del suolo. Questo potrebbe implicare che i sistemi agrivoltaici sono più resilienti ai cambiamenti climatici rispetto alle monoculture. Un esempio di sistema agrivoltaico è mostrato nella Figura 1.1



Figura 1.1. Impianto agrivoltaico (Fonte: euractiv.it)

1.3 Storia

Il concetto di agrivoltaico è stato proposto per la prima volta da Adolf Goetzberger e Armin Zastrow [3], due fisici tedeschi, presso l'istituto Fraunhofer in Germania, nel 1982. Essi proposero di elevare i pannelli fotovoltaici a circa due metri dal suolo e di aumentare la distanza tra le file, in modo tale da ottenere una radiazione uniforme sul terreno e consentire allo stesso tempo il passaggio di attrezzature agricole. L'alto costo della tecnologia fotovoltaica, tuttavia, consentì di predisporre prototipi ed impianti agrivoltaici sperimentali solo all'inizio del nuovo millennio.

Nel 2004 l'ingegnere giapponese Akira Nagashima [14] sviluppò il primo sistema agrivoltaico, la cui struttura era simile ad un pergolato da giardino. Egli progettò diversi campi di prova con tassi di ombreggiamento differenti basati sul concetto del punto di saturazione di ciascuna coltura, le quali impiegano solo una piccola percentuale della radiazione solare incidente (tra il 3% e il 6%) per raggiungere il loro massimo tasso di fotosintesi. L'idea era di condividere la restante radiazione solare con i pannelli fotovoltaici per generare elettricità.

Il primo modello pilota sperimentale, tuttavia, fu installato in Francia, vicino alla città di Montpellier nel 2010. Questo sistema consisteva in moduli fotovoltaici monocristallini montati su palafitte larghe 0.8 m, ad un'altezza di 4 m e inclinati con un angolo di 25 gradi. Al fine di valutare l'effetto di ombreggiamento creato dai pannelli il prototipo venne diviso in due parti a diverse densità: la prima, chiamata "*full density*", aveva una spaziatura ottimale tra le file di pannelli, che trasmetteva in media il 50% della radiazione solare alle colture. L'altra, chiamata "*half density*", invece, è stata realizzata tramite l'eliminazione di una fila di pannelli ogni due, lasciando così, in media, il 70% della radiazione solare incidente a disposizione delle colture.

Negli ultimi anni, sono stati attuati diversi progetti agrivoltaici in tutto il mondo. In Germania, l'istituto Fraunhofer per i sistemi a energia solare (ISE) è all'avanguardia in questo campo. Nel 2016 gli scienziati di questo istituto hanno realizzato a Heggelbach, vicino al lago di Costanza in Germania, un impianto sperimentale agrivoltaico sotto il progetto "APV-RESOLA" [13], con l'intento di esaminare gli impatti di questa tecnologia per quanto riguarda aspetti quali la produzione di energia, la produzione agricola, la fattibilità economica e l'accettazione sociale. L'impianto ha una potenza di 194 kWp e occupa una superficie di 0.3 ettari, con pannelli montati a cinque metri dal terreno su una struttura sopraelevata. Inoltre, sono stati realizzati dal medesimo istituto altri tre impianti sperimentali vicino Santiago del Cile, per investigare i loro impatti sulle colture da campo in regioni aride e ad alta radiazione solare.

Nello stesso anno un impianto di ricerca su piccola scala venne installato in Arizona ad opera del gruppo di ricerca di Barron-Gafford, concentrandosi sulla crescita di colture comuni nelle terre aride come i peperoni e i pomodori.

In Italia, l'azienda REM Tec S.r.l., insieme all'Università di Piacenza, hanno creato il sistema di inseguimento solare chiamato Agrovoltaico® [2]. I primi due impianti sono stati installati nel 2012 a Castelvetro Piacentino e Monticelli d'Ongina, coprendo un'area rispettivamente di 7 e 20 ettari. Questa tecnologia consiste in un sistema biassiale, composto da un asse principale orizzontale su cui

sono presenti degli assi secondari che sostengono i pannelli, i quali possono ruotare grazie ad un motore elettrico. Il sistema è posto a 5 m di altezza da terra, in grado di massimizzare la produzione di energia elettrica, lasciando il terreno libero per l'agricoltura. Una rappresentazione del sistema è visibile in Figura 1.2.

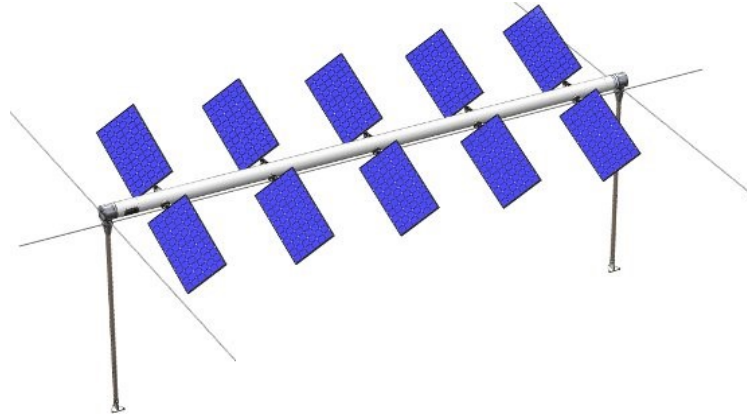


Figura 1.2. Rappresentazione Sistema Agrivoltaico (Fonte: remtec.energy)

Globalmente, i sistemi agrivoltaici sono in continua crescita: secondo una stima, dal 2014 sono stati installati più di 2000 impianti in tutto il mondo, con la maggior parte di essi situati in Giappone. La capacità complessiva installata al gennaio 2020 era pari a circa 2.8 GWp.

1.4 Descrizione dell'impianto

Poiché un impianto agrivoltaico si basa sulla combinazione di produzione di energia fotovoltaica e produzione agricola, esso è composto da pannelli fotovoltaici tradizionali sopraelevati.

Il fotovoltaico è una forma di tecnologia per ricavare energia rinnovabile, basata sulla conversione dell'energia della luce solare in energia elettrica o calore attraverso l'uso di pannelli fotovoltaici. Il pannello è il cuore dell'intero sistema perché permette la cattura della luce e, attraverso le celle che lo compongono, rende possibile il trasferimento di energia. Esistono diversi tipi di sistemi fotovoltaici, la loro classificazione si basa sul posizionamento o sulla struttura delle celle.

La produzione di energia da parte dei pannelli fotovoltaici si basa sul cosiddetto "effetto fotovoltaico".

L'effetto fotovoltaico, scoperto nel 1839 dal fisico francese Alexandre Edmond Becquerel, è un fenomeno fisico che consiste nella conversione diretta della radiazione solare in una corrente continua di elettroni. Le celle fotovoltaiche convenzionali consistono nella giunzione di due sottili strati di materiale semiconduttore, uno strato è detto "di tipo P", quindi di tipo positivo, l'altro è detto "di tipo N", di tipo negativo. Il materiale di tipo P è composto generalmente da silicio a cui vengono aggiunte impurità di altri metalli, come il boro, in modo che esso possieda un deficit di elettroni liberi, chiamate lacune. Al contrario, il materiale di tipo N è composto da silicio con aggiunte di metalli, come il

fosforo, che permettano al materiale finale di possedere un surplus di elettroni liberi. L'attività di inserimento di altri atomi all'interno del silicio è chiamata drogaggio. Collegando questi due strati di materiali, si ottiene la cosiddetta "giunzione p-n" in cui si ha la creazione di un campo elettrico, a causa dell'attrazione di alcuni elettroni presenti nel lato n-type verso l'altro materiale positivo. Quando i fotoni con un'appropriata lunghezza d'onda, derivanti dalla radiazione solare, colpiscono la giunzione p-n, possono trasferire la loro energia ad alcuni elettroni liberi, facendoli quindi passare a uno stato eccitato e rendendoli liberi di muoversi. Il movimento interessa anche le lacune, che essendo per definizione un'assenza di elettroni, si muovono in senso opposto ad essi. In questo modo, si ha la creazione di corrente elettrica nella cella.

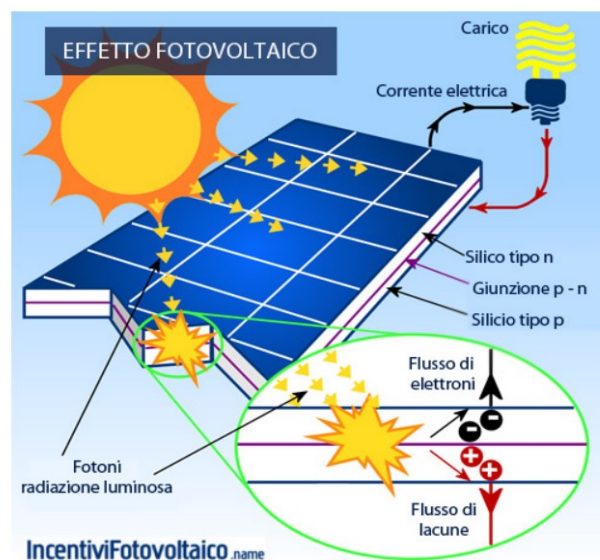


Figura 1.3. Funzionamento pannello fotovoltaico (Fonte: led-italia.it)

Un impianto fotovoltaico è costituito essenzialmente da:

- Un generatore
- Un sistema di condizionamento e controllo della potenza
- Un accumulatore di energia

La componente elementare di un generatore fotovoltaico è rappresentata dalla cella. Essa è costituita da una sottile lamina di un materiale semiconduttore, molto spesso silicio. È qui che si verifica la conversione della radiazione solare in corrente elettrica, secondo l'effetto fotovoltaico descritto precedentemente. Una cella fotovoltaica è generalmente di forma quadrata, ma può essere anche rotonda, con una superficie compresa tra i 100 e i 225 cm² e uno spessore di circa 0.3 mm. Essa si comporta come una minuscola batteria, producendo, nelle condizioni di soleggiamento tipiche dell'Italia, una corrente di 3 A, con una tensione di 0.5 V, quindi una potenza di 1,5 W.

A seconda dei loro processi di produzione si distinguono i seguenti tipi di celle fotovoltaiche:

- Celle monocristalline: vengono prodotte tagliando una barra monocristallina. Il vantaggio di questa tipologia è dato dall'alto rendimento, fino al 16%. Questo tipo di celle è però molto

costoso a causa del complicato processo di produzione. Sono caratterizzate di solito da un'omogenea colorazione blu.

- Celle policristalline: sono, invece, generate dagli scarti del taglio dei lingotti monocristallini. Il rendimento è inferiore rispetto alle prime, intorno il 10-12%, però sono più economiche.
- Celle amorphe: vengono realizzate tramite deposizione catodica di atomi di silicio su una piastra di vetro. Questo tipo di cella ha un rendimento minore, attorno al 4%, ma si adatta in caso di irraggiamento diffuso. Le celle così prodotte sono riconoscibili dal caratteristico colore scuro e possono essere realizzate secondo qualsiasi forma geometrica.



Figura 1.4. Tipologie di celle fotovoltaiche: monocristalline, policristalline e amorphe (Fonte:Qookka.com)

Le celle fotovoltaiche vengono poi assemblate in modo opportuno a costituire un'unica struttura, il modulo fotovoltaico. I moduli più diffusi sono costituiti da 36 celle disposte su quattro file parallele e sono in grado di raggiungere, ad oggi, la potenza massima di 165 Wp.

Più moduli collegati in serie formano un pannello fotovoltaico, ovvero una struttura ancorabile al suolo o agli edifici, e analogamente più pannelli insieme formano una stringa. Infine, l'unione di diverse stringhe, generalmente in parallelo per fornire la potenza richiesta, costituisce il generatore fotovoltaico.

La potenza massima del generatore fotovoltaico è determinata dalla somma delle potenze di picco di ciascun componente, misurate alle condizioni standard (STC, *Standard Test Conditions*)

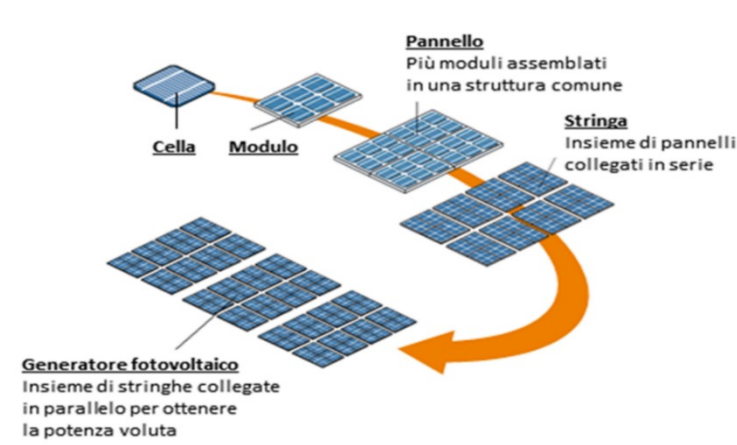


Figura 1.5. Generatore fotovoltaico (Fonte: www.adec-impianti.it)

Questo tipo di impianto presenta diversi vantaggi: non solo consente la produzione di energia senza provocare alcun tipo di inquinamento, ma inoltre non richiede frequente manutenzione, è scarsamente soggetto ad usura e consente di realizzare l'indipendenza energetica dai tradizionali fornitori di energia.

Il sistema di condizionamento e controllo della potenza è costituito da un inverter, dispositivo che ha il compito di trasformare la corrente continua prodotta dai moduli in corrente alternata, da un trasformatore e da un sistema di rifasamento e filtraggio che garantisce la qualità della potenza in uscita. Trasformatore e sistema di filtraggio sono normalmente inseriti all'interno dell'inverter.

1.5 Analisi della letteratura

Viste le potenzialità dei sistemi agrivoltaici, molti ricercatori hanno concentrato i loro studi sull'argomento. I calcoli da loro effettuati hanno dimostrato che l'applicazione di questi sistemi può aumentare i redditi delle aziende agricole di oltre il 30%, se le perdite nella resa vengono minimizzate dalla selezione di colture adatte.

Dupraz e Marrou [5], ricercatori dell'INRA (Institut national de la recherche agronomique), hanno condotto degli esperimenti, tra il 2011 e il 2013, con l'obiettivo di valutare se il tasso di crescita delle colture viene influenzato dall'ombra dei pannelli e se le variabili climatiche sono significativamente modificate dalla stessa. In particolare, sono state monitorate le condizioni climatiche al di sotto dei pannelli fotovoltaici, come la temperatura del suolo, la temperatura e l'umidità dell'aria, la velocità del vento e le radiazioni incidenti, a intervalli regolari di tempo. Gli esperimenti sono stati condotti confrontando diverse soluzioni impiantistiche: FD (*full density*) e HD (*half density*), come descritte precedentemente, e in condizioni di pieno sole FS (*full sun*), durante tre stagioni meteorologiche (inverno, primavera ed estate). Al di sotto dei pannelli sono state piantate colture a ciclo breve, come lattuga e cetriolo, e colture a ciclo lungo, quale il grano duro.

Per la temperatura media, l'umidità dell'aria e la velocità del vento i valori registrati sono risultati simili nelle diverse condizioni, indipendentemente dalla stagione climatica. Invece, la temperatura del suolo è diminuita in modo significativo sotto i pannelli, di circa 2°C, ed è stata riscontrata una riduzione dell'escursione termica giornaliera delle colture poiché la temperatura media diurna è diminuita mentre quella notturna è aumentata. Il tasso di crescita delle colture è rimasto pressoché invariato. La lattuga ha dimostrato un'elevata adattabilità all'ombreggiamento causato dall'impianto, con un incremento dell'area totale delle foglie ombreggiate ma una riduzione del numero delle stesse. I cetrioli hanno mostrato uno sviluppo ridotto solo durante la fase giovanile, tre settimane dopo la semina, che potrebbe derivare da variazioni nella temperatura del suolo, ma un inalterato sviluppo durante il resto della loro crescita. Il grano duro, al di sotto dell'impianto fotovoltaico, ha mostrato una maturazione in ritardo di circa tre giorni.

Con questo studio Dupraz e Marrou sono stati quindi in grado di dimostrare che i rendimenti dell'esperimento agrivoltaico sono superiori rispetto alle tradizionali monoculture, grazie all'uso del

Land Equivalent Ratio (LER) [5]. Questo è un indicatore della produttività del terreno utilizzato di solito per valutare la resa dei sistemi misti, con più colture sulla stessa superficie. Il concetto di Land Equivalent Ratio può essere esteso a qualsiasi sistema che combina due tipi di produzione sulla stessa unità di terreno, quindi anche per l'agrivoltaico. In tal caso, esso è definito come la somma di due rapporti secondo la Formula 1.1:

$$LER = \frac{Resa\ agricola_{APV}}{Resa\ agricola_{no\ PV}} + \frac{Produzione\ elettricità_{APV}}{Produzione\ elettricità_{solo\ PV}} \quad (1.1)$$

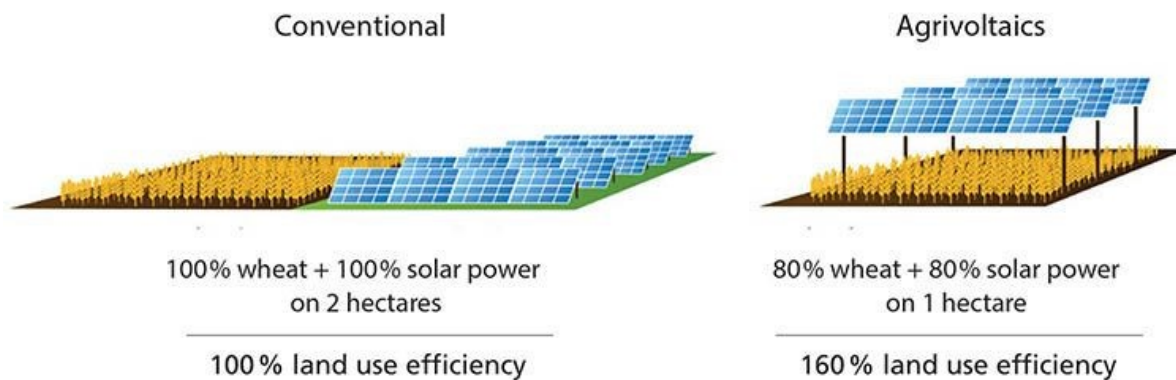


Figura 1.6. Rappresentazione del Land Equivalent Ratio (Fonte: www.infobuildenergia.it.)

Teoricamente se non vi sono differenze tra i sistemi colturali l'indice sarà pari a 1. Quando il LER è maggiore di 1 indica la presenza di una interazione positiva tra le specie consociate e di conseguenza un aumento della produttività rispetto alle monoculture. I valori ottenuti dagli esperimenti condotti sono stati di 1.35 e 1.73 rispettivamente per FD e HD: ciò implica che sarebbe necessario il 35%-73% di terreno aggiuntivo al fine di produrre la stessa quantità di energia e biomassa, su due superfici distinte.

Alcuni anni dopo, Dinesh e Pierce della Michigan Technological University, negli Stati Uniti, hanno condotto uno studio volto a realizzare un'analisi tecnico-economica di un impianto agrivoltaico [6]. A tal fine sono stati utilizzati due software: PVsyst attraverso cui è stata effettuata un'analisi di sensitività per analizzare la produzione di energia per unità di superficie variando l'angolo di inclinazione, l'efficienza di conversione e la spaziatura tra le file dei pannelli e STICS per ottenere dei dati sulla resa delle colture.

Anche per questo esperimento sono state adottate le stesse configurazioni di Dupraz e Marrou, FD e HD. Per entrambe le tipologie i moduli sono stati montati ad un'altezza di 4 m dal suolo, con un angolo di inclinazione di 25°. Nella configurazione *half density* la distanza dei pannelli era di 6.4 m, mentre per quella *full density* di 3.2 m. Al di sotto dei pannelli è stata coltivata la lattuga, poiché dagli studi condotti precedentemente è risultata essere una coltura tollerante all'ombra e quindi maggiormente compatibile con i sistemi agrivoltaici. Dai risultati è emerso che la soluzione *half density* presenta l'efficienza media annua più alta rispetto alle altre due configurazioni, *full density* e

a piena luce. Per la lattuga è stata osservata una riduzione della resa del 42% in FD e del 19% in HD durante i mesi estivi rispetto alle coltivazioni in pieno sole. In primavera, invece, è stata registrata una riduzione del 21% per FD, mentre per la tipologia HD non si sono registrate variazioni significative nei raccolti. Ciò era dovuto alle condizioni di ombreggiamento moderate durante la semina primaverile. Questo combinato alla capacità di adattamento della lattuga all'ombreggiamento hanno fatto sì che i rendimenti rimanessero praticamente inalterati.

In un recente studio, Schindele, del *Fraunhofer Institute for Solar Energy Systems* (ISE), e altri ricercatori, hanno riscontrato che la coltivazione di grano e patate negli impianti agrivoltaici è risultata negativa o positiva a seconda delle condizioni atmosferiche.

Tuttavia, gli studi sopra riportati non hanno analizzato gli impatti dell'agrivoltaico sulla condizione idrica del suolo, a differenza di quello svolto da Amaducci, docente dell'Università Cattolica del Sacro Cuore di Piacenza, e altri studiosi.

Essi hanno analizzato la resa del mais alimentato a pioggia sotto il sistema Agrovoltaico®, citato precedentemente. La simulazione è stata condotta utilizzando un set di dati climatici di 40 anni da una città del Nord Italia e diverse configurazioni di agrivoltaico, le quali differiscono in base alla densità del pannello e all'impostazione dell'inseguimento solare. La riduzione dell'irraggiamento globale sotto il sistema Agrovoltaico® è stata maggiormente influenzata dalla densità dei pannelli, (29.5% e 13.4% rispettivamente per doppia e singola densità) piuttosto che dalla gestione degli stessi (23,2% e 20,0% rispettivamente per i pannelli solari e statici).

Inoltre, dai risultati è emerso che quando l'acqua è un fattore non limitante, la resa del mais sotto il sistema è leggermente inferiore a quella ottenibile in condizioni di piena luce. Tuttavia, in condizioni di stress da siccità, la resa media del mais è risultata più alta e stabile.

Infine, Amaducci e gli altri studiosi hanno concluso che è necessario effettuare un'analisi economica e ambientale per fornire una valutazione completa della sostenibilità del sistema Agrovoltaico®, in modo tale da determinare e ottimizzarne le prestazioni ambientali ed economiche. [2]

CAPITOLO 2

STRUMENTI STATISTICI

2.1 Modelli di regressione per serie storiche

2.1.1 Le serie storiche

Una serie storica è definita come una successione di dati numerici nella quale ogni dato è associato ad un particolare istante od intervallo del tempo [4].

Le serie storiche possono essere

- univariate: se in ciascun punto o intervallo del tempo viene osservato un unico fenomeno
- multivariate o multiple: se le variabili osservate per ogni istante sono più d'una
- equispaziate: se le osservazioni vengono rilevate ad intervalli di tempo equidistanti tra loro

In una serie storica l'ordine dei dati non è accidentale; infatti, la loro caratteristica peculiare è data dalla dipendenza tra osservazioni successive.

Le serie storiche che si usano nella pratica provengono da ambiti diversi e per questo ne esistono di diverse tipologie

- serie storiche economiche: scaturiscono dai fenomeni economici (es: serie storica del PIL)
- serie storiche fisiche: nascono dallo studio di fenomeni astronomici, fisici, metereologici (es: livello delle precipitazioni, delle temperature)
- serie storiche demografiche: derivano dallo studio di fenomeni demografici quali nascite, morti, immigrazione, popolazione residente
- serie storiche generate da controllo di processo: ottenute dalla misurazione ripetuta di una variabile legata al funzionamento di un dato processo
- serie storiche da processi binari: si hanno quando la serie può assumere solamente due valori, di solito indicati con 0 e 1
- serie storiche generate da processi di punto: si hanno quando si considera una successione di eventi che avvengono ad intervalli di tempo casuali

Una serie storica, ottenuta osservando il fenomeno Y , viene generalmente indicata secondo la seguente notazione: y_1, y_2, \dots, y_n dove y_t indica il valore della variabile Y al tempo t .

Il processo generatore dei dati di una serie storica può essere rappresentato dal Modello 2.1:

$$Y_t = f(t) + u_t \quad (2.1)$$

dove

- $f(t)$ è una sequenza completamente deterministica che rappresenta la parte sistematica della serie

- u_t è una sequenza di variabili casuali che rappresenta la parte stocastica della serie e obbedisce ad una data legge di probabilità

Entrambe le componenti non sono osservabili, ma vanno determinate sulla base del campione.

2.1.2 Approccio classico vs moderno

L'approccio classico alle serie storiche si concentra sulla determinazione della "legge di evoluzione temporale" del fenomeno, che viene rappresentata dalla componente $f(t)$, mentre la componente u_t , puramente casuale, non viene considerata esplicitamente in Y_t .

Quest'ultima si ipotizza che sia generata da un processo *white noise*, ovvero da una successione di variabili casuali, indipendenti e identicamente distribuite, con media nulla e varianza costante, indicato come segue:

$$\varepsilon_t \sim \text{WN}(0, \sigma^2_\varepsilon)$$

- $E(\varepsilon_t) = 0 \quad \forall t$
- $\text{Var}(\varepsilon_t) = \sigma^2 \quad \forall t$
- $E(\varepsilon_r, \varepsilon_s) = 0 \quad \forall r, s, r \neq s$

Una successione di variabili casuali per cui si abbia $\text{Cov}(u_r, u_s) = 0 \quad \forall r, s, r \neq s$ viene detta processo stocastico a componenti incorrelate.

Nell'approccio moderno, invece, si suppone che la serie storica sia già stata depurata dalla sua parte deterministica e ci si concentra solo sullo studio della parte stocastica, che si ipotizza essere un processo a componenti correlate del tipo: $u_t = g(Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t$

2.1.3 Le componenti di una serie storica

Nelle serie storiche si assume che la parte sistematica $f(t)$ risulti dall'azione congiunta di tre componenti, non direttamente osservabili, che sono il trend, il ciclo e la stagionalità. [4]

Il trend

Detto anche componente tendenziale, è la tendenza di fondo del fenomeno considerato, riferita ad un lungo periodo di tempo. Esso mette in evidenza lo sviluppo e l'evoluzione strutturale del sistema economico, che si manifestano lentamente nel tempo. Di solito si ipotizza che i valori del trend siano esprimibili mediante una funzione. A volte si impone alla funzione la caratteristica di monotonia, che permette di considerare la tendenza del fenomeno come crescente o decrescente. Le funzioni più comunemente utilizzate sono quelle lineari ed esponenziali.

Il ciclo

Il ciclo o componente congiunturale è originato dalla presenza di fluttuazioni attribuite al succedersi di fasi ascendenti e discendenti del fenomeno considerato, solitamente collegate alle fasi di espansione e di recessione del sistema economico. Con fluttuazioni si intende gli scostamenti verso l'alto o verso il basso dei dati rispetto al trend di natura più o meno regolare.

La stagionalità

È una componente periodica i cui effetti si esauriscono nel corso dell'anno. In particolare, è costituita dai movimenti del fenomeno nel corso dell'anno che tendono a ripetersi in maniera pressappoco identica nello stesso periodo. Questo avviene a causa dell'influenza di fattori climatici e sociali, quali l'alternarsi delle stagioni, le variazioni climatiche, ma anche tradizioni o convenzioni tipiche dell'organizzazione sociale.

Le tre componenti possono combinarsi principalmente secondo tre modelli:

a) Il modello additivo: $Y_t = T_t + C_t + S_t + \varepsilon_t$

Le componenti sono indipendenti ed espresse tutte nella stessa unità di misura

b) Il modello moltiplicativo: $Y_t = T_t * C_t * S_t * \varepsilon_t$

La componente T_t è espressa nella stessa unità di misura di Y_t mentre le altre componenti assumono la forma di numeri puri. Vi è dipendenza tra le componenti. Il modello moltiplicativo può essere ricondotto a quello additivo tramite trasformata logaritmica

$$\log Y_t = \log T_t + \log C_t + \log S_t + \log \varepsilon_t$$

c) Il modello misto: $Y_t = T_t * C_t * S_t + \varepsilon_t$

Le componenti T_t e ε_t sono espresse nella stessa unità di misura di Y_t , mentre C_t e S_t assumono la forma di numeri indici.

2.1.4 Modelli di regressione lineare

L'obiettivo dei modelli di regressione è quello di capire se e come il comportamento di una variabile di interesse può essere spiegato usando altre variabili.

La variabile di interesse, indicata con Y , ha il ruolo di variabile risposta e viene anche chiamata variabile dipendente o spiegata. Vi sono poi altre variabili concomitanti, chiamate X_1, \dots, X_p , che hanno il ruolo di variabili esplicative. Queste vengono anche chiamate predittori, regressori, variabili esogene, variabili indipendenti o covariate. Si vuole determinare come la prima viene influenzata dalle seconde [8].

Il modello di regressione è formato da due componenti legate in maniera additiva: una componente deterministica, ossia non casuale, $g(x_{i1}, \dots, x_{ip}, \theta)$ e una componente casuale ε_i . La relazione tra la risposta e le esplicative viene spiegata dalla componente deterministica, che rappresenta la parte strutturale di tale relazione. Il termine casuale comprende, invece, l'insieme di circostanze che non vengono spiegate completamente dalla componente deterministica.

Nel modello di regressione lineare la relazione tra la variabile dipendente e i regressori è appunto di tipo lineare. Essa può essere semplice o multipla, a seconda del fatto che il numero di esplicative sia una o più.

Il modello di regressione lineare semplice

È il modello di regressione più semplice, poiché le variabili coinvolte sono solamente due: la variabile risposta Y e una variabile esplicativa, indicata con X.

In particolare, la relazione tra le due è espressa dalla Formula 2.2:

$$Y_i = \beta_1 + \beta_2 * x_i + \epsilon_i, \quad (2.2)$$

dove β_1 e β_2 rappresentano i parametri di regressione.

Le assunzioni su cui si basa tale modello sono

- i. Assunzione di linearità
- ii. Assunzione di omoschedasticità

L'assunzione di linearità si riferisce ai parametri e non alle variabili. Quella di omoschedasticità implica che la varianza non cambia al variare di i, ossia la capacità della retta di rappresentare il fenomeno non varia al variare della grandezza considerata.

I parametri del modello possono essere stimati utilizzando un metodo chiamato dei minimi quadrati. Lo stimatore dei minimi quadrati sceglie i coefficienti di regressione in modo tale che la retta stimata sia il più possibile vicina ai dati osservati. Si ottengono le seguenti stime per i parametri

$$\widehat{\beta}_2 = \frac{\text{COV}(X,Y)}{\text{var}(X)}$$
$$\widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 * \bar{x}$$

Il modello di regressione lineare multipla

La regressione lineare multipla rappresenta un'estensione di quella semplice. In questo modello vengono considerate più variabili esplicative e si vuole spiegare il loro effetto congiunto sulla variabile risposta.

Il modello può essere rappresentato come segue: $Y_i = \beta_1 * x_{i1} + \beta_2 * x_{i2} + \dots + \beta_p * x_{ip} + \epsilon_i$

In questo caso le assunzioni su cui si basa tale modello sono

- i. Linearità
- ii. Media nulla, omoschedasticità, normalità e indipendenza degli errori
- iii. Indipendenza lineare tra le variabili esplicative

Esso può venire anche rappresentato in forma matriciale secondo la formula 2.3

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_i \\ \dots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{i1} & \dots & x_{ip} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_i \\ \dots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \dots \\ \epsilon_i \\ \dots \\ \epsilon_n \end{bmatrix} \quad (2.3)$$
$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon}$$

La matrice X, detta matrice delle condizioni sperimentali o matrice di regressione, è stocastica e a rango pieno, ossia le sue colonne sono linearmente indipendenti.

La stima del vettore dei parametri si ottiene secondo la seguente formula: $\hat{\beta} = (X^T X)^{-1} X^T y$

2.1.5 Il coefficiente di determinazione R^2

È un indice che misura la bontà di adattamento ai dati del modello di regressione lineare. Esso è calcolato tramite il rapporto tra la devianza spiegata SQ_{reg} , cioè la somma dei quadrati dovuti alla regressione, e la devianza totale SQ_{tot} , data dalla somma dei quadrati totali.

$$R^2 = \frac{SQ_{reg}}{SQ_{tot}} = 1 - \frac{SQ_{res}}{SQ_{tot}} = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (2.4)$$

L'indice R^2 misura la capacità delle variabili esplicative di spiegare, attraverso il modello lineare, il comportamento della risposta. Questo coefficiente può assumere dei valori compresi tra 0 e 1. Più il valore dell' R^2 è alto e più il modello si adatta bene ai dati.

Nel modello di regressione lineare semplice l' R^2 coincide con l'indice di correlazione al quadrato.

2.1.6 Stima del trend mediante funzioni polinomiali

La componente di trend può venire rappresentata tramite una qualche funzione del tempo, che deve essere stimata. Per semplificare si ipotizza che la parte sistematica della serie $f(t)$ sia composta da solo trend, il quale si assume essere un polinomio di grado q del tipo: $f(t) = \alpha_0 + \alpha_1 * t + \dots + \alpha_q * t^q$.

Il modello di regressione che ne deriva

$$y_t = \alpha_0 + \alpha_1 * t + \dots + \alpha_q * t^q + \varepsilon_t \quad (2.5)$$

è lineare nei parametri e facilmente stimabile. Esso può essere espresso tramite forma matriciale nel seguente modo: $\mathbf{y} = \mathbf{P}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$ dove \mathbf{y} , $\boldsymbol{\varepsilon}$ sono dei vettori di lunghezza n (numerosità della serie), $\boldsymbol{\alpha}$ un vettore di lunghezza q e \mathbf{P} è la matrice di regressione.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 1 & 1 & 1^2 & \dots & 1^q \\ 1 & 2 & 2^2 & \dots & 2^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & n & n^2 & \dots & n^q \end{bmatrix} \quad (2.6)$$

Tramite il principio dei minimi quadrati si ottengono le stime dei parametri: $\hat{\boldsymbol{\alpha}} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{y}$.

L'ordine q del polinomio dipende dal comportamento di fondo della serie storica. I casi più comuni sono i seguenti:

- $Y_t = \alpha_0 + \varepsilon_t$ trend costante

- $Y_t = \alpha_0 + \alpha_1 * t + \varepsilon_t$ trend lineare
- $Y_t = \alpha_0 + \alpha_1 * t + \alpha_2 * t^2 + \varepsilon_t$ trend parabolico
- $Y_t = \alpha_0 + \alpha_1 * t + \alpha_2 * t^2 + \alpha_3 * t^3 + \varepsilon_t$ trend cubico

Utilizzando un grado del polinomio abbastanza alto si riesce ad ottenere un accostamento stretto tra i valori stimati e quelli osservati; tuttavia, il modello perde gradi di libertà all'aumentare di q. Per la scelta di q si può ricorrere al confronto tra i coefficienti di determinazione R^2 .

2.1.7 Il confronto tra R^2 per la scelta del grado del polinomio

Nel modello di regressione lineare l'aggiunta di una nuova variabile esplicativa causa una riduzione della varianza residua e un conseguente aumento del valore dell'indice R^2 . Indicato con R^2_r il coefficiente di determinazione calcolato tramite stima della funzione di regressione polinomiale di grado r, si ha sempre che $R^2_r \leq R^2_{r+1}$. Invece questo può non accadere se si usa la versione "corretta" di tale coefficiente $\overline{R^2}$. Il criterio per la scelta del grado del polinomio si fonda sul confronto tra le versioni corrette degli R^2 . In particolare, si sceglie un polinomio di grado r se $\overline{R^2}_r \geq \overline{R^2}_{r+1}$, mentre si prosegue nella ricerca per valori di r superiori se $\overline{R^2}_r < \overline{R^2}_{r+1}$.

Il coefficiente di determinazione R^2 corretto è dato dalla seguente formula:

$$R_a^2 = 1 - \left[\frac{n-1}{n-k-1} \right] * (1 - R^2) \quad (2.7)$$

dove n rappresenta il numero di osservazioni e k il numero di variabili indipendenti.

2.2 Regressione quantilica

La regressione quantilica è un tipo di regressione capace di individuare più effetti rispetto alle procedure convenzionali. La regressione classica si focalizza sul valore atteso di una variabile Y condizionata ai valori di un gruppo di variabili X, $E(Y|X)$, la cosiddetta funzione di regressione. Questa funzione può essere più o meno complessa, ma è ristretta esclusivamente a una specifica posizione della distribuzione condizionata Y.

La regressione quantilica (QR) estende questo approccio, permettendo di studiare la distribuzione condizionata di Y su X a diverse posizioni, offrendo, così, una visione globale delle interrelazioni tra Y e X. In particolare, modella la relazione tra un insieme di variabili indipendenti e i quantili specifici di una variabile dipendente, il più delle volte la mediana [3]. Inoltre, presenta due vantaggi principali rispetto alla regressione dei minimi quadrati ordinari:

- Non fa ipotesi sulla distribuzione della variabile dipendente
- Tende a resistere all'influenza di osservazioni anomale

Venne introdotta da Koenker e Basset (1978) come un'estensione della stima ai minimi quadrati.

Sia Y una generica variabile casuale. La sua media è definita come centro della distribuzione che minimizza la somma al quadrato delle deviazioni, e si trova dalla soluzione del seguente problema di minimizzazione: $\mu = \operatorname{argmin}_c E(Y-c)^2$.

La mediana invece minimizza la somma assoluta delle deviazioni, ossia: $Me = \operatorname{argmin}_c E|Y-c|$
 Usando le osservazioni del campione, possiamo trovare gli stimatori $\hat{\mu}$ e \widehat{Me} .

I quantili sono definiti come particolari posizioni della distribuzione, ossia il θ -esimo quantile è il valore y tale per cui $P(Y \leq y) = \theta$.

La funzione quantile è definita come l'inverso della funzione di ripartizione, in questo modo:

$$Q_Y(\theta) = Q(\theta) = F^{-1}_Y(\theta) = \inf\{y: F(y) > \theta\} \text{ per } 0 < \theta < 1. \quad (2.8)$$

I quantili sono definiti anche minimizzando la somma assoluta pesata delle deviazioni. In questo caso il quantile è: $q_\theta = \operatorname{argmin}_c E[\rho_\theta(Y-c)]$, dove $\rho_\theta = [\theta - I(y < 0)]y = [(1-\theta)I(y \leq 0) + \theta I(y > 0)]y$ è una funzione di perdita. Questa è una funzione di perdita asimmetrica assoluta, data dalla somma pesata delle deviazioni assolute, con peso $(1-\theta)$ per le deviazioni negative e peso θ a quelle positive.

Nel caso di una variabile discreta Y con distribuzione di probabilità $f(y)=P(Y=y)$, il precedente problema di minimizzazione diventa:

$$\begin{aligned} q_\theta &= \operatorname{argmin}_c E[\rho_\theta(Y - c)] \\ &= \operatorname{argmin}_c \left\{ (1 - \theta) \sum_{y \leq c} |y - c|f(y) + \theta \sum_{y > c} |y - c|f(y) \right\}. \end{aligned} \quad (2.9)$$

Lo stesso criterio viene adottato nel caso di una variabile continua, sostituendo le sommatorie con gli integrali:

$$\begin{aligned} q_\theta &= \operatorname{argmin}_c E[\rho_\theta(Y - c)] \\ &= \operatorname{argmin}_c \left\{ (1 - \theta) \int_{-\infty}^c |y - c|f(y)d(y) + \theta \int_c^{+\infty} |y - c|f(y)d(y) \right\} \end{aligned} \quad (2.10)$$

Dove $f(y)$ rappresenta la funzione di densità di probabilità di Y . Per $\theta=0,5$ si ottiene la mediana, definita con la formula sopra.

Questo problema può essere generalizzato tramite la regressione. Infatti, interpretando Y come una variabile risposta e X come un set di variabili predittive, l'idea di media incondizionata come minimizzatore dell'equazione $\mu = \operatorname{argmin}_c E(Y-c)^2$ può essere estesa alla stima di una funzione di media condizionata:

$$\hat{\mu}(\mathbf{x}_i, \boldsymbol{\beta}) = \operatorname{argmin}_\mu E[Y - \mu(\mathbf{x}_i, \boldsymbol{\beta})]^2, \quad (2.11)$$

dove $\mu(x_i, \beta) = E[Y|X = x_i]$ è la funzione di media condizionata. Nel caso di una funzione di media lineare, la precedente equazione diventa:

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} E[Y - \mathbf{x}_i^\top \beta]^2 \quad (2.12)$$

ottenendo così il modello di regressione lineare dei minimi quadrati.

Lo stesso approccio può essere usato per un generico quantile θ -esimo. In questo caso si ottiene:

$$\hat{q}_Y(\theta, \mathbf{X}) = \underset{Q_Y(\theta, \mathbf{X})}{\operatorname{argmin}} E[\rho_\theta(Y - Q_Y(\theta, \mathbf{X}))], \quad (2.13)$$

dove $Q_Y(\theta, X) = Q_\theta[Y|X = x]$ rappresenta la generica funzione dei quantili condizionata. Allo stesso modo, per il modello lineare la precedente equazione diventa:

$$\hat{\beta}(\theta) = \underset{\beta}{\operatorname{argmin}} E[\rho_\theta(Y - \mathbf{X}\beta)], \quad (2.14)$$

2.2.1 Il modello QR più semplice: Il caso del regressore dummy

La forma più semplice di modello QR prevede una variabile risposta quantitativa e una variabile esplicativa di tipo dummy. Questa semplice impostazione mira a studiare le differenze nella variabile risposta tra i due gruppi determinati dalla variabile esplicativa dicotomica. La stima del modello diventa: $\hat{Y}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x_{i1}$, dove Y è la variabile risposta e x l'esplicativa di tipo dummy che assume valore 1 in corrispondenza di un dato gruppo e valore 0 per un altro. Per diversi valori di $\theta \in [0, 1]$ permette di ottenere uno stimatore dei quantili della variabile risposta per i due gruppi considerati. Usando la codifica gruppo 1 = 0 e gruppo 2 = 1 per la variabile dicotomica si avrà:

$$\begin{aligned} \hat{Y}_\theta &= \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)*0 = \hat{\beta}_0(\theta) \text{ per il gruppo 1} \\ \hat{Y}_\theta &= \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)*1 = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta) \text{ per il gruppo 2} \end{aligned} \quad (2.15)$$

Dunque, $\hat{\beta}_0(\theta)$ fornisce la stima del θ -esimo quantile condizionato della variabile risposta Y per il gruppo 1 mentre per il secondo gruppo, la stima del θ -esimo quantile è data dalla somma dei parametri $\hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)$.

Variando θ nell'intervallo $[0,1]$ il gruppo di intercette stimate offre una stima della distribuzione della variabile risposta nel primo gruppo. Per il secondo, la Y è ottenuta tramite la somma del gruppo di intercette e coefficienti angolari per diversi θ .

2.2.2 Modello QR con regressore nominale

Il semplice modello descritto sopra può essere leggermente complicato sostituendo la variabile dicotomica con una categoriale a g categorie. In questo caso la regressione quantilica permette di confrontare i differenti quantili tra i g diversi gruppi. È noto che per gestire variabili nominali a g livelli è necessario introdurre nel modello $g-1$ variabili dummy.

Il modello QR che ne deriva è del tipo:

$$\hat{Y}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)*I(x_{i1}) + \dots + \hat{\beta}_{g-1}(\theta)*I(x_{ig-1}), \quad (2.16)$$

dove $I(\cdot)$ è la funzione indicatrice che dà valore 1 se la variabile assume il valore tra parentesi e 0 altrimenti. Il primo livello, così detto livello di riferimento, è associato all'intercetta.

Questa equazione permette di ottenere lo stimatore del θ -esimo quantile per i diversi livelli del regressore.

Per il livello di riferimento associato all'intercetta, il modello si riduce a: $\hat{Y}_\theta = \hat{\beta}_0(\theta)$

Per il secondo livello si ottiene: $\hat{Y}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)$

Infine, per il g -esimo livello si ottiene: $\hat{Y}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_{g-1}(\theta)$

Quindi, in un modello con un regressore nominale, per un dato quantile θ , la combinazione dell'intercetta con diverse pendenze, consente di stimare il quantile condizionato della variabile risposta. L'effetto stimato del particolare gruppo è ottenuto usando la variabile dummy associata con quella particolare pendenza.

2.2.3 Modello QR con regressore quantitativo

In questo caso il modello è del tipo: $\hat{Y}_\theta = \hat{\beta}_0(\theta) + \hat{\beta}_1(\theta)x_{i1}$ dove x_{i1} è una variabile quantitativa. I parametri stimati nei modelli di regressione quantilica lineare hanno la stessa interpretazione di quelli stimati tramite altri modelli lineari. In modo analogo al modello OLS, il coefficiente $\beta_i(\theta)$ del modello QR può essere interpretato come il tasso di variazione del quantile θ della distribuzione della variabile dipendente per unità di variazione nel valore dell' i -esimo regressore:

$$\beta_i(\theta) = \frac{\partial Q_\theta(Y|\mathbf{X})}{\partial x_i}. \quad (2.17)$$

2.3 ANOVA e ANCOVA

L'analisi della varianza (ANOVA, dall'inglese Analysis of Variance) è una tecnica utilizzata per testare le differenze nei valori medi di un numero qualsiasi di gruppi. L'obiettivo dell'ANOVA è quello di valutare gli effetti su una variabile di interesse, variabile risposta di tipo continuo, di uno o

più fattori di controllo, variabili indipendenti categoriali con due o più modalità. Essa si divide in varie tipologie a seconda di quante e come sono le variabili dipendenti e indipendenti.

2.3.1 ANOVA a un fattore

L'analisi della varianza ad un fattore, o a una via, viene usata quando si ha un unico fattore e si vuole verificare se diversi livelli di tale fattore abbiano un effetto misurabile sulla variabile dipendente.

Indicando con y_{ij} l' i -esima osservazione nel j -esimo gruppo, con $i = 1, \dots, n_j$ indice dell'osservazione per gruppo e $j = 1, \dots, J$ indice del gruppo, si calcolano quindi le medie dei j gruppi $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$, le varianze campionarie corrette $s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ij} - \bar{y}_j)^2$ e la media generale $\bar{y} = \frac{1}{n} \sum_{j=1}^J \sum_{i=1}^{n_j} y_{ij} = \frac{1}{n} \sum_{j=1}^J n_j \bar{y}_j$, dove $n = \sum_{j=1}^J n_j$

I calcoli dell'ANOVA a un fattore consistono nella scomposizione della devianza totale presente nei dati osservati in devianza dovuta alle differenze tra i gruppi (devianza tra i gruppi) e devianza dovuta alle differenze interne ai gruppi (devianza entro i gruppi), secondo la Formula 2.18:

$$\begin{aligned} \sum_{j=1}^J \sum_{i=1}^{n_j} (y_{ij} - \bar{y})^2 &= \sum_{j=1}^J (n_j - 1) s_j^2 + \sum_{j=1}^J n_j (\bar{y}_j - \bar{y})^2 \\ \text{Dev. totale} &= \text{Dev. entro i gruppi} + \text{Dev. tra i gruppi} \quad (2.18) \\ SQ_{tot} &= SQ_{res} + SQ_{reg} \end{aligned}$$

La somma totale dei quadrati (SQ_{tot}) può essere composta nella somma dei quadrati all'interno dei gruppi (SQ_{res}) e nella somma dei quadrati tra i gruppi (SQ_{reg}). Se la variabilità tra i gruppi è più elevata rispetto a quella interna ai gruppi, cioè se $SQ_{reg} > SQ_{res}$ allora vi sono delle differenze fra le medie dei gruppi. Invece se risulta l'inverso allora le medie dei gruppi sono approssimativamente uguali.

Per l'ANOVA ad un fattore si assume che le osservazioni siano realizzazioni di variabili casuali normali indipendenti, ossia $Y_{ij} \sim N(\mu_j, \delta^2)$

Essa può essere vista come un test d'ipotesi, in cui si vuole verificare l'ipotesi nulla di omogeneità delle medie (H_0) contro l'alternativa che non tutte le medie siano uguali (H_1)

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_j \\ H_1: \text{le } \mu_i \text{ non sono tutte uguali tra loro} \end{cases} \quad (2.19)$$

La statistica test per questa verifica d'ipotesi viene data da

$$F = \frac{SQ_{reg} / (J-1)}{SQ_{res} / (n-J)} \sim F_{J-1, n-J} \text{ sotto } H_0 \quad (2.20)$$

Fonte di variazione	Gradi di libertà	Somma dei quadrati (devianza)	Varianza	F^{OSS}	α^{OSS}
Tra gruppi	$J - 1$	$SQ_{reg} = \sum_j n_j (\bar{y}_j - \bar{y})^2$	$\frac{\sum_j n_j (\bar{y}_j - \bar{y})^2}{J-1}$	$\frac{\sum_j n_j (\bar{y}_j - \bar{y})^2}{\frac{\sum_{i,j} (y_{ij} - \bar{y}_j)^2}{n-J}}$	$P(F_{J-1, n-J} > F^{OSS})$
Entro gruppi	$n - J$	$SQ_{res} = \sum_{i,j} (y_{ij} - \bar{y}_j)^2$	$\frac{\sum_{i,j} (y_{ij} - \bar{y}_j)^2}{n-J}$		

Tabella 2.1 ANOVA ad un fattore

L'ipotesi nulla viene rifiutata se il valore osservato della statistica F è maggiore del quantile di livello $1-\alpha$ della distribuzione di riferimento, per un dato α . In questo caso il test è significativo, ovvero esiste almeno una significativa differenza tra le medie nei gruppi.

I risultati vengono solitamente sintetizzati e riassunti in una tabella, detta tabella dell'ANOVA, come riportato sotto

L'ANOVA ad un fattore può essere formulata anche attraverso un modello di regressione lineare multipla. Di conseguenza, il test F riportato sopra può essere interpretato come un test di nullità di un gruppo di coefficienti di regressione del modello.

Il modello può essere scritto come: $Y_i = \beta_1 + \beta_2 * x_{i2} + \dots + \beta_J * x_{iJ} + \varepsilon_i$ con $\varepsilon_i \sim N(0, \delta^2)$, dove si è posto $x_{i1} = 1 \forall i$ e con $x_{ij} \quad j = 1, \dots, (J-1)$, si indicano le variabili esplicative indicatrici, che rappresentano i diversi livelli del fattore. Nel caso di J gruppi è necessario ricorrere a J-1 variabili indicatrici.

L'ipotesi nulla di uguaglianza delle medie è dunque formulabile come

$$H_0 = \beta_2 = \dots = \beta_J = 0 \quad (2.21)$$

La statistica F diventa quindi:

$$F = \frac{(\tilde{\sigma}^2 - \hat{\sigma}^2)/(J - 1)}{\hat{\sigma}^2/(n - J)} \stackrel{H_0}{\sim} F_{J-1, n-J}. \quad (2.22)$$

dove $\tilde{\sigma}^2$ è la stima di massima verosimiglianza della varianza nel modello nullo, calcolata come $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{SQ_{tot}}{n}$ e $\hat{\sigma}^2$ è la stima di massima verosimiglianza nel modello completo. Si dimostra che tale statistica coincide con quella riportata precedentemente.

2.3.2 ANOVA a due fattori

Il modello dell'ANOVA a un fattore può essere esteso al caso in cui le osservazioni sulla variabile risposta vengono classificate secondo due fattori. Si ha così il modello dell'ANOVA a due fattori.

Il suo obiettivo è investigare l'effetto dei due fattori sulla variabile risposta, ossia si vuole definire se la distribuzione della variabile dipendente rimane la stessa per i vari livelli dei fattori. Quest'ultimi possono avere un effetto separato oppure interagire. A seconda che i fattori interagiscano o meno tra loro si avranno due tipi di ANOVA.

ANOVA a due fattori senza interazione

In questo modello si inseriscono, per ogni fattore, tante variabili indicatrici quanti sono i livelli di tale fattore escluso il primo che viene preso a riferimento. Quindi se il primo fattore presenta J livelli e il secondo fattore è diviso in K livelli, il numero totale di parametri di regressione è $(J-1)+(K-1)+1 = J+K-1$. Posto $x_{i1} = 1 \forall i$ il modello risulta: $Y_i = \mu + \alpha_2 * x_{i2} + .. + \alpha_j * x_{ij} + \gamma_2 * x_{i,J+1} + .. + \gamma_K * x_{i,J+K-1} + \epsilon_i$ con $\epsilon_i \sim N(0, \delta^2)$ indipendenti $i = 1, \dots, n$, e dove x_{i2}, \dots, x_{ij} rappresentano le variabili indicatrici riferite al primo fattore e $x_{i,J+1}, \dots, x_{i,J+K-1}$ quelle riferite al secondo fattore.

Anche in questo caso la devianza totale (SQ_{tot}) può essere scomposta in una parte dovuta al primo fattore (SQ_I), una dovuta al secondo (SQ_{II}) e una parte residua (SQ_{res})

Per l'assenza dell'effetto del primo fattore si ha l'ipotesi nulla $H_0: \alpha_2 = \dots = \alpha_j = 0$ e la statistica test usata per valutare l'effetto del primo fattore a J livelli è

$$F = \frac{SQ_I / (J - 1)}{SQ_{res} / (n - (J + K - 1))} \stackrel{H_0}{\sim} F_{J-1, n-(J+K-1)} \quad (2.23)$$

Analogamente per l'assenza di effetto del secondo fattore l'ipotesi nulla è del tipo $H_0: \gamma_2 = \dots = \gamma_k = 0$ e la distribuzione della statistica test di riferimento è $F_{K-1, n-(J+K-1)}$.

Invece per l'assenza degli effetti di entrambi i fattori l'ipotesi nulla risulta $H_0: \alpha_2 = \dots = \alpha_j = \gamma_2 = \dots = \gamma_k = 0$ e la distribuzione è $F_{J+K-2, n-(J+K-1)}$.

Di seguito viene riportata la Tabella 1.2 dell'ANOVA a due fattori senza interazione

Fonte	G.d.L.	Devianza	Varianza	F^{oss}	Valore-p
I fattore	$J - 1$	SQ_I	$\frac{SQ_I}{J-1}$	$\frac{\frac{SQ_I}{J-1}}{\frac{SQ_{res}}{n-(J+K-1)}}$	$P(F_{J-1, n-(J+K-1)} > F^{oss})$
II fattore	$K - 1$	SQ_{II}	$\frac{SQ_{II}}{K-1}$	$\frac{\frac{SQ_{II}}{K-1}}{\frac{SQ_{res}}{n-(J+K-1)}}$	$P(F_{K-1, n-(J+K-1)} > F^{oss})$
Residua	$n - (J + K - 1)$	SQ_{res}	$\frac{SQ_{res}}{n-(J+K-1)}$		
Totale	$n - 1$	SQ_{tot}	$\frac{SQ_{tot}}{n-1}$		

Tabella 1.2 ANOVA a due fattori senza interazione

ANOVA a due fattori con interazione

Questo modello viene costruito utilizzando le stesse variabili indicatrici definite per il modello senza interazione. L'interazione viene modellata tramite i prodotti delle variabili indicatrici dei due fattori. Il modello diventa quindi: $Y_i = \alpha_2 * x_{i2} + .. + \alpha_j * x_{ij} + \gamma_2 * x_{i,J+1} + .. + \gamma_K * x_{i,J+K-1} + \delta_1 * x_{i2} * x_{i,J+1} + \dots + \epsilon_i$ con $\epsilon_i \sim N(0, \delta^2)$ indipendenti $i = 1, \dots, n$. Complessivamente si avranno $J * K$ coefficienti, tanti quanti sono i gruppi. Tale modello consente di verificare anche l'ipotesi di assenza di interazione in termini di nullità dei coefficienti: $H_0: \delta_1 = \dots = \delta_{(J-1)(K-1)}$

La statistica F per la significatività di una delle fonti di variabilità, primo fattore, secondo fattore o interazione, può essere espressa in questo modo

$$F = \frac{SQ(\text{fonte})/GdL(\text{fonte})}{SQ_{res}/(n - JK)} \stackrel{H_0}{\sim} F_{GdL(\text{fonte}),n-JK} \quad (2.24)$$

2.3.3 ANCOVA

L'analisi della varianza è un modello lineare che permette di usare solo variabili esplicative qualitative. Esistono però delle situazioni in cui, tra le variabili esplicative, possono essere presenti sia variabili qualitative che quantitative. In questo caso il modello adatto da utilizzare prende il nome di ANCOVA, o analisi della covarianza.

Il modello può essere rappresentato come segue: $Y_i = \beta_1 * x_{i1} + \beta_2 * x_{i2} + \beta_3 * x_{i3} + \beta_4 * x_{i4} + \epsilon_i$, dove x_{i2} rappresenta la variabile quantitativa, x_{i3} la variabile qualitativa e x_{i4} l'interazione tra le due.

Per verificare l'esistenza di un effetto di interazione tra le due variabili l'ipotesi necessaria sarà quella di nullità del coefficiente β_4

CAPITOLO 3

APPLICAZIONE AI DATI

Nel seguente capitolo verranno analizzati 3 diversi dataset, contenenti rilevazioni riguardanti la crescita di tre diverse tipologie di coltivazioni: erba medica, fagioli e patate. I prossimi paragrafi analizzeranno separatamente le varie tipologie di piante secondo i primi due modelli descritti nel capitolo precedente.

3.1 Coltura erba medica

3.1.1 Descrizione dei dati

Il dataset riporta i valori di alcuni parametri di piante di erba medica fatte crescere in diverse condizioni di luce. È costituito da 7 colonne e 1350 righe. Le colonne indicano rispettivamente:

- **date:** data in cui sono stati rilevati i valori dei parametri delle diverse piante. I parametri delle piante sono stati rilevati ogni settimana in un periodo compreso tra il 22 aprile 2021 e il 29 luglio 2021.
- **week:** settimana in cui sono stati rilevati i parametri delle diverse piante. In totale sono state considerate 15 settimane.
- **light:** condizione di luce o ombra in cui è cresciuta la pianta
 - FL (full light): in pieno sole
 - SH (shadow): sotto l'ombra dei pannelli
- **area:** area di riferimento. Ciascuna condizione di luce e ombra è stata divisa in più aree. La parte FL è stata divisa in 6 aree diverse, mentre la parte SH è stata divisa in 24 aree diverse. In totale sono stati studiati 30 settori. La rappresentazione delle aree è mostrata in Figura 3.1

FL		SH			
1FL	2FL	1SH	2SH	3SH	4SH
3FL	4FL	5SH	6SH	7SH	8SH
5FL	6FL	9SH	10SH	11SH	12SH
		13SH	14SH	15SH	16SH
		17SH	18SH	19SH	20SH
		21SH	22SH	23SH	24SH

Figura 3.1. Rappresentazione della suddivisione di FL e SH in aree

- plant: numero della pianta. In ciascuna area sono state coltivate tre piante, per un numero complessivo di 90.
- height: altezza delle piante. Primo parametro rilevato.
- number_leaves: numero delle foglie delle piante. Secondo parametro rilevato.

I grafici dell'andamento dell'altezza nel corso delle settimane, per ciascuna area considerata, vengono riportati di seguito. Quelli relativi al numero di foglie vengono riportati in Appendice.

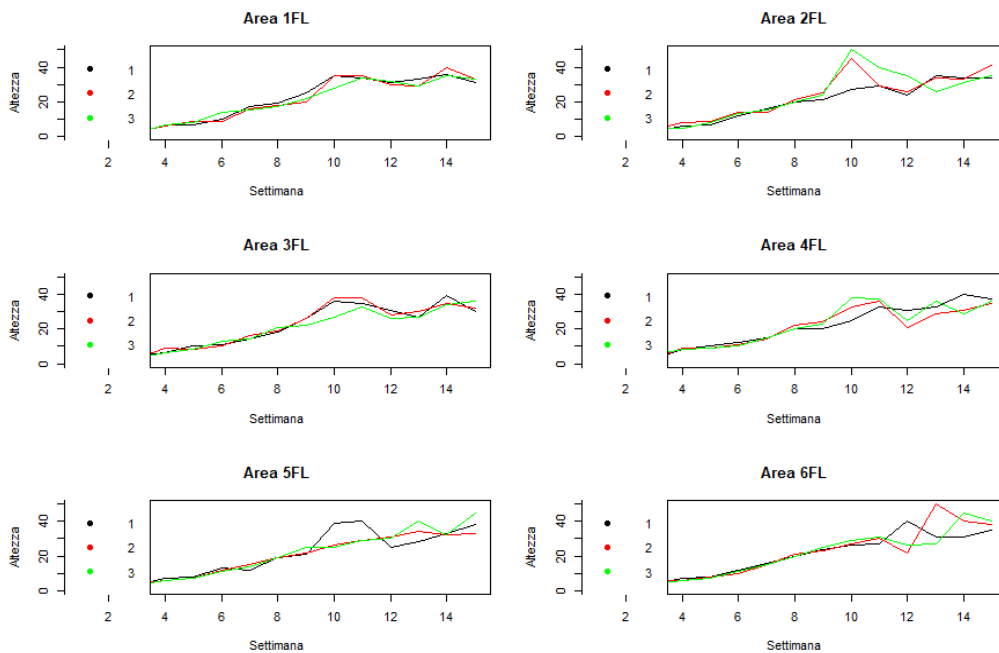


Figura 3.2. Grafici dell'andamento dell'altezza per erba medica per le aree FL

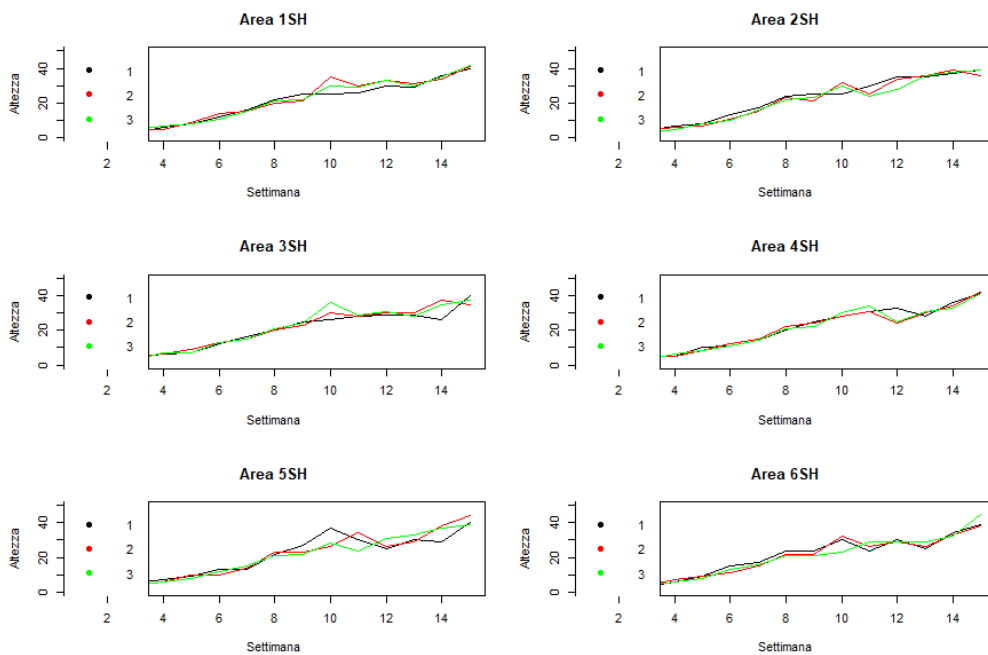


Figura 3.3. Grafici dell'andamento dell'altezza per erba medica per le aree 1SH-6SH

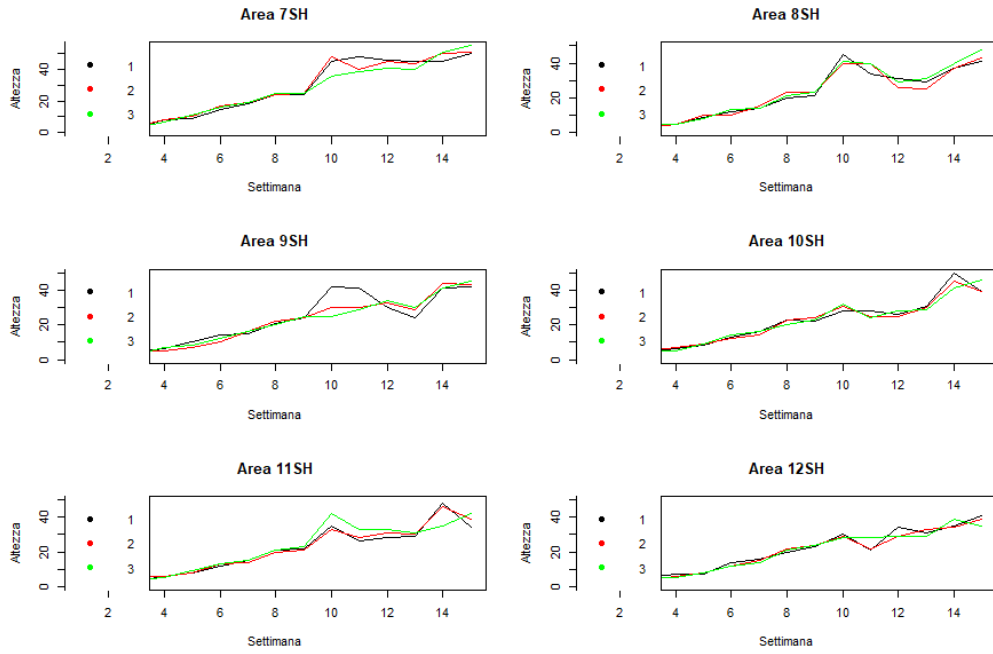


Figura 3.4. Grafici dell'andamento dell'altezza per erba medica per le aree 7SH-12SH

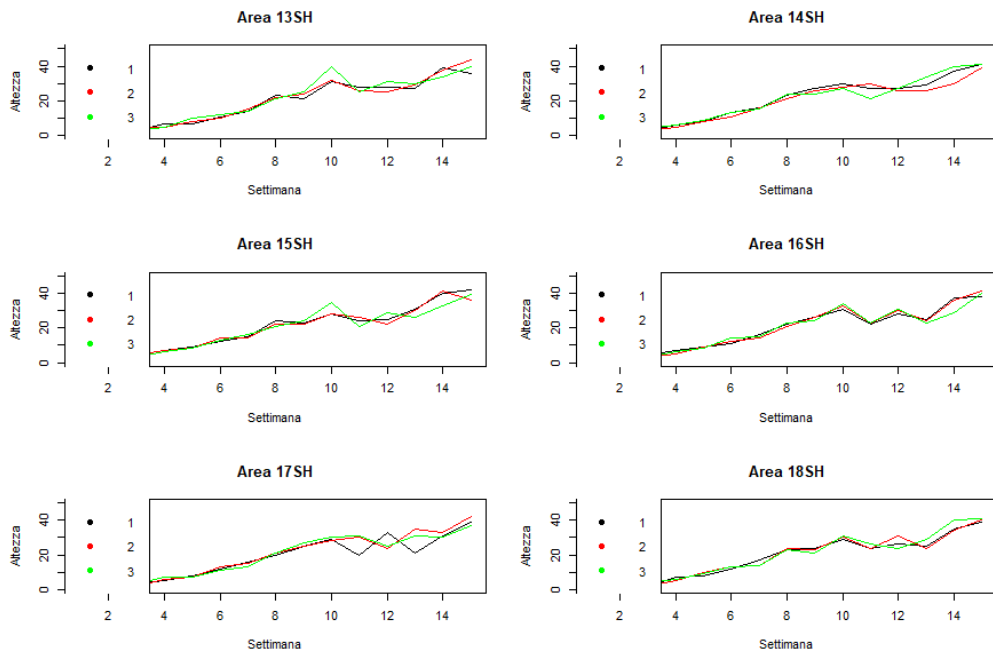


Figura 3.5. Grafici dell'andamento dell'altezza per erba medica per le aree 13SH-18SH

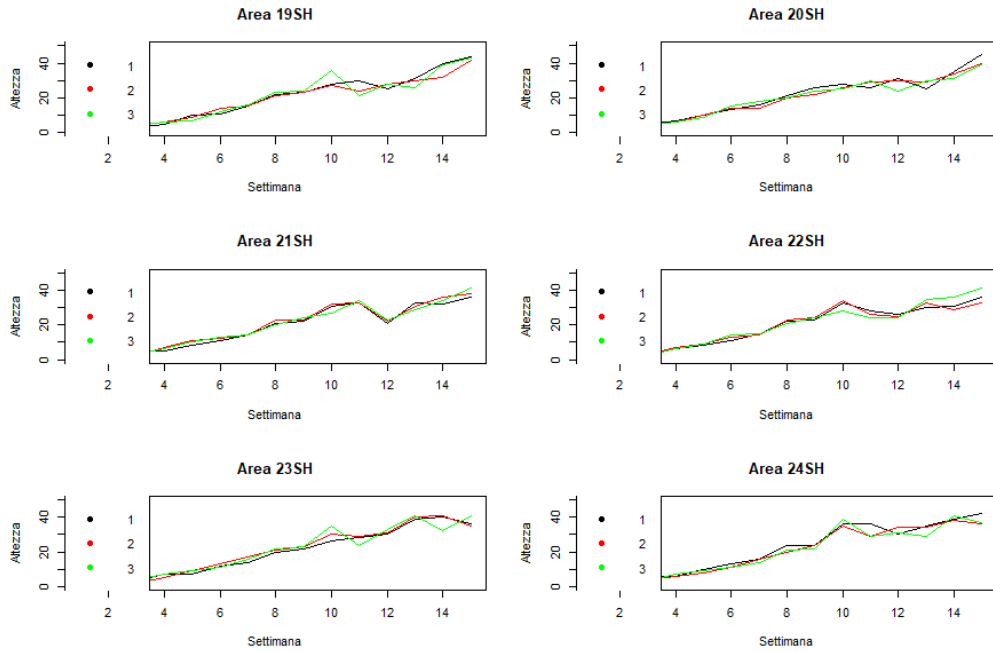


Figura 3.6. Grafici dell'andamento dell'altezza per erba medica per le aree 19SH-24SH

3.1.2 Modelli di regressione con trend polinomiale

Nei modelli stimati nel presente paragrafo, verranno considerate come variabili risposta “height” e “number_leaves” e come regressori “light” e “week”, poiché è ragionevole supporre che l’altezza e il numero di foglie della pianta dipendano dalle condizioni di luce e dal trascorrere del tempo. Poiché la variabile “light” è di tipo qualitativo, per inserirla nel modello si è creata una variabile dummy così definita:

$$x_{il} = d_l = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{se light} = \text{SH} \end{cases} \quad (3.1)$$

Per la stima si è preso in considerazione il modello con trend cubico e il termine di interazione tra le esplicative.

Il modello con variabile risposta “height” ha prodotto i seguenti risultati:

```
Call:
lm(formula = altezza ~ d1 + tempo + tempo2 + tempo3 + d1:tempo +
    d1:tempo2 + d1:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-13.5301  -1.7302  -0.2594   1.3071  22.5404
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.742023   0.614741  -2.834 0.004669 **
d1           2.991026   1.374603   2.176 0.029735 *
tempo       1.156182   0.322011   3.591 0.000342 ***
tempo2      0.291476   0.045994   6.337 3.19e-10 ***
tempo3     -0.012753   0.001894  -6.734 2.44e-11 ***
d1:tempo    -2.274584   0.720038  -3.159 0.001619 **
d1:tempo2    0.388017   0.102847   3.773 0.000168 ***
```

d1:tempo3 -0.017802 0.004235 -4.204 2.80e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.846 on 1342 degrees of freedom
 Multiple R-squared: 0.9166, Adjusted R-squared: 0.9162
 F-statistic: 2108 on 7 and 1342 DF, p-value: < 2.2e-16

Il modello stimato risulta: $\hat{y}_i = -1.742023 + 2.991026*d1 + 1.156182*tempo + 0.291476*tempo^2 - 0.012753*tempo^3 - 2.274584*d1*tempo + 0.388017*d1*tempo^2 - 0.012753*d1*tempo^3$.

Considerando le due condizioni di luce e ombra separatamente il modello diventa

$\hat{y}_i = (-1.742023 + 2.991026) + (1.156182 - 2.274584)*tempo + (0.291476 + 0.388017)*tempo^2 + (-0.012753 - 0.012753)*tempo^3$ per FL

$\hat{y}_i = -1.742023 + 1.156182*tempo + 0.291476*tempo^2 - 0.012753*tempo^3$ per SH

Dunque, risulta

$$\hat{y}_i = \begin{cases} 1.249003 - 1.118402 * tempo + 0.679493 * tempo^2 - 0.025506 * tempo^3 & \text{per FL} \\ -1.742023 + 1.156182 * tempo + 0.291476 * tempo^2 - 0.012753 * tempo^3 & \text{per SH} \end{cases} \quad (3.2)$$

I coefficienti dei termini di interazione tra le due esplicative risultano tutti significativi, portando così al rifiuto dell'ipotesi di nullità dei coefficienti. Dunque, l'effetto della variabile temporale cambia a seconda dei valori della variabile dummy, ossia è diverso per FL e SH.

Il valore del parametro β_1 indica che l'altezza media delle piante è maggiore di 2.991026 per la condizione FL rispetto a SH.

Il coefficiente di determinazione R^2 risulta pari a 0.9166, cioè il 91.66% della variabilità della variabile risposta altezza viene spiegata dal modello.

Il grafico di dispersione tra l'altezza e il tempo, con le curve di regressione stimate, viene riportato di seguito:

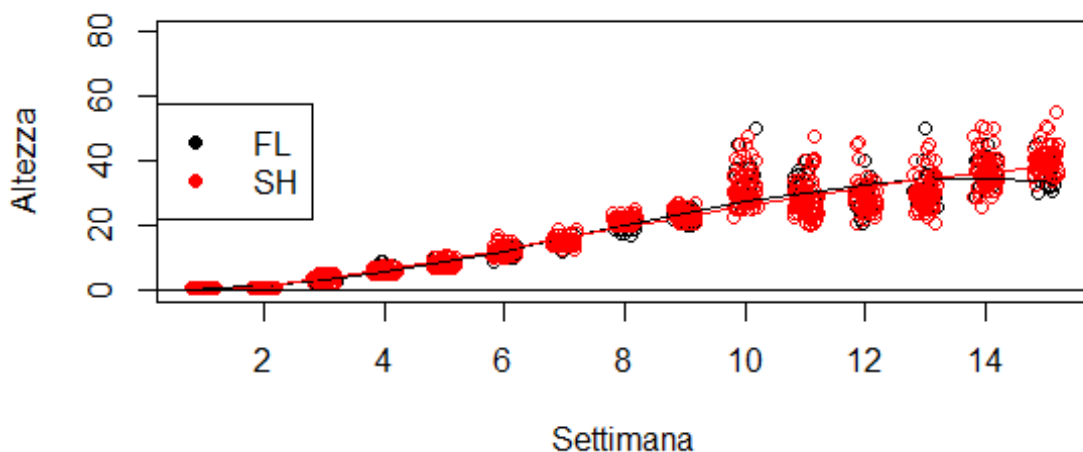


Figura 3.7. Diagramma di dispersione tra altezza e tempo per erba medica e curve di regressione stimate con trend cubico

Dal grafico si può notare che le due curve sono approssimativamente coincidenti; tuttavia, dalla settimana 13 iniziano a divergere. Sarebbe quindi necessario disporre di dati in un arco temporale più esteso per migliorare l'analisi.

Il modello con variabile risposta “number_leaves” ha prodotto i seguenti risultati:

```
Call:
lm(formula = nfoglie ~ d1 + tempo + tempo2 + tempo3 + d1:tempo +
    d1:tempo2 + d1:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-11.7076  -1.4008  -0.4008   1.5469  23.2735
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.004792   0.522178  -3.839 0.000129 ***
d1           2.081675   1.167626   1.783 0.074841 .
tempo       1.096196   0.273525   4.008 6.47e-05 ***
tempo2      0.382897   0.039069   9.801 < 2e-16 ***
tempo3     -0.021182   0.001609  -13.168 < 2e-16 ***
d1:tempo   -1.781195   0.611620  -2.912 0.003648 **
d1:tempo2  0.336173   0.087361   3.848 0.000125 ***
d1:tempo3  -0.015225   0.003597  -4.233 2.47e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 3.267 on 1342 degrees of freedom
Multiple R-squared:  0.9202, Adjusted R-squared:  0.9198
F-statistic: 2211 on 7 and 1342 DF, p-value: < 2.2e-16
```

Il modello stimato è il seguente: $\hat{y}_i = -2.004792 + 2.081675 \cdot d1 + 1.096196 \cdot tempo + 0.382897 \cdot tempo^2 - 0.021182 \cdot tempo^3 - 1.781195 \cdot d1 \cdot tempo + 0.336173 \cdot d1 \cdot tempo^2 - 0.015225 \cdot d1 \cdot tempo^3$

Se consideriamo FL e SH separatamente il modello diventa:

$\hat{y}_i = (-2.004792 + 2.081675) + (1.096196 - 1.781195) \cdot tempo + (0.382897 + 0.336173) \cdot tempo^2 + (-0.021182 - 0.015225) \cdot tempo^3$ per FL

$\hat{y}_i = -2.004792 + 1.096196 \cdot tempo + 0.382897 \cdot tempo^2 - 0.021182 \cdot tempo^3$ per SH

Quindi si può riscrivere:

$$\hat{y}_i = \begin{cases} 0.076883 - 0.684999 \cdot tempo + 0.71907 \cdot tempo^2 - 0.036407 \cdot tempo^3 & \text{per FL} \\ -2.004792 + 1.096196 \cdot tempo + 0.382897 \cdot tempo^2 - 0.021182 \cdot tempo^3 & \text{per SH} \end{cases} \quad (3.3)$$

Anche in questo caso i termini di interazione risultano tutti significativi. Dunque, la stima del trend è differente nelle due condizioni FL e SH.

Il coefficiente di determinazione R^2 è pari a 0.9202, ossia il modello spiega il 92.02% della variabilità della risposta. Il grafico delle curve stimate dal modello è riportato nella Figura 3.8.

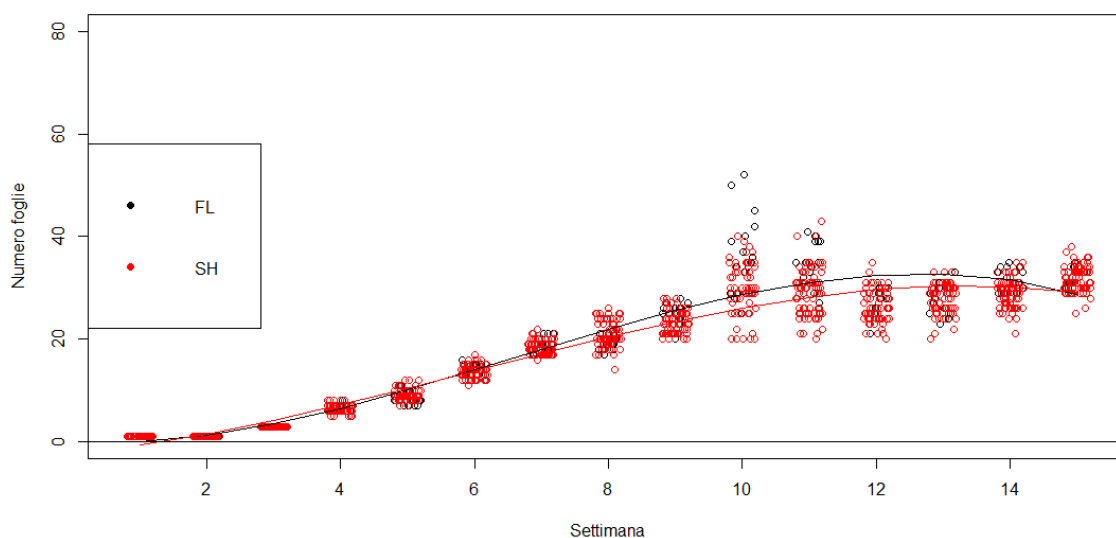


Figura 3.8. Diagramma di dispersione tra numero di foglie e tempo per erba medica e curve di regressione stimate con trend cubico

3.1.3 Regressione quantilica

In questi modelli, analogamente alla regressione classica, verranno considerate come variabili risposta “height” e “number_leaves” e invece come regressori “light” e “week”. Poiché la variabile “light” è di tipo qualitativo, per inserirla nel modello si è creata una variabile dummy così definita:

$$xi1 = d1 = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{se light} = \text{SH} \end{cases} \quad (3.4)$$

Verranno analizzati, in particolare, i quantili 0.10 e 0.90.

Il quantile di ordine 0.10 analizza la dinamica temporale per le piante che presentano valori bassi della variabile risposta.

Il quantile di ordine 0.90 analizza la dinamica temporale per le piante che presentano valori alti della variabile risposta.

I modelli con variabile risposta altezza hanno prodotto i seguenti risultati:

```
Call: rq(formula = altezza ~ tempo + tempo2 + tempo3 + d1 + d1:tempo +
d1:tempo2 + d1:tempo3, tau = taus, data = erba.medica)
```

```
tau: [1] 0.1
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-2.02667	0.61492	-3.29580	0.00101
tempo	1.13556	0.47435	2.39392	0.01681
tempo2	0.20667	0.09546	2.16500	0.03056
tempo3	-0.00889	0.00496	-1.79350	0.07312
d1	1.12923	0.94979	1.18892	0.23468
tempo:d1	-0.79581	0.73044	-1.08950	0.27613
tempo2:d1	0.12702	0.14484	0.87699	0.38065
tempo3:d1	-0.00571	0.00731	-0.78119	0.43483

```
Call: rq(formula = altezza ~ tempo + tempo2 + tempo3 + d1 + d1:tempo +
d1:tempo2 + d1:tempo3, tau = taus, data = erba.medica)
```

```
tau: [1] 0.9
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.10967	0.31599	3.51176	0.00046
tempo	-0.75176	0.24946	-3.01356	0.00263
tempo2	0.67076	0.04616	14.53223	0.00000
tempo3	-0.02867	0.00212	-13.54877	0.00000
d1	0.51371	0.67353	0.76271	0.44577
tempo:d1	-0.69954	0.54330	-1.28759	0.19811
tempo2:d1	0.19721	0.11200	1.76082	0.07850
tempo3:d1	-0.01138	0.00562	-2.02271	0.04330

Per il quantile di ordine 0.10, ossia per piante con valori bassi della variabile risposta altezza, risultano significativi i coefficienti relativi all'intercetta, al tempo e al tempo² e debolmente significativo quello relativo a tempo³, mentre non risulta significativo nessun termine di interazione. Dunque, per queste piante non vi è differenza significativa tra la condizione FL e SH.

Per il quantile di ordine 0.90, ossia per piante con valori alti della variabile risposta altezza, risultano significativi i coefficienti relativi all'intercetta, al tempo e alle sue potenze, e debolmente significativi i termini di interazione d1×tempo² e d1×tempo³.

Ricalcolando i modelli eliminando le variabili che non risultano significative si ottengono i seguenti risultati:

```
Call: rq(formula = altezza ~ tempo, tau = 0.1)
```

```
tau: [1] 0.1
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-4.50000	0.08555	-52.59925	0.00000
tempo	2.50000	0.02824	88.53151	0.00000

```
Call: rq(formula = altezza ~ tempo + tempo2 + tempo3 + d1:tempo2 +
d1:tempo3, tau = 0.9)
```

```
tau: [1] 0.9
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.14286	0.38663	2.95592	0.00317
tempo	-0.79545	0.31047	-2.56207	0.01051
tempo2	0.68182	0.06333	10.76606	0.00000
tempo3	-0.02922	0.00320	-9.12053	0.00000
tempo2:d1	0.09169	0.04236	2.16442	0.03061
tempo3:d1	-0.00730	0.00299	-2.44217	0.01473

Le curve di regressione quantilica ottenute sono riportate nel grafico riportato nella Figura 3.9:

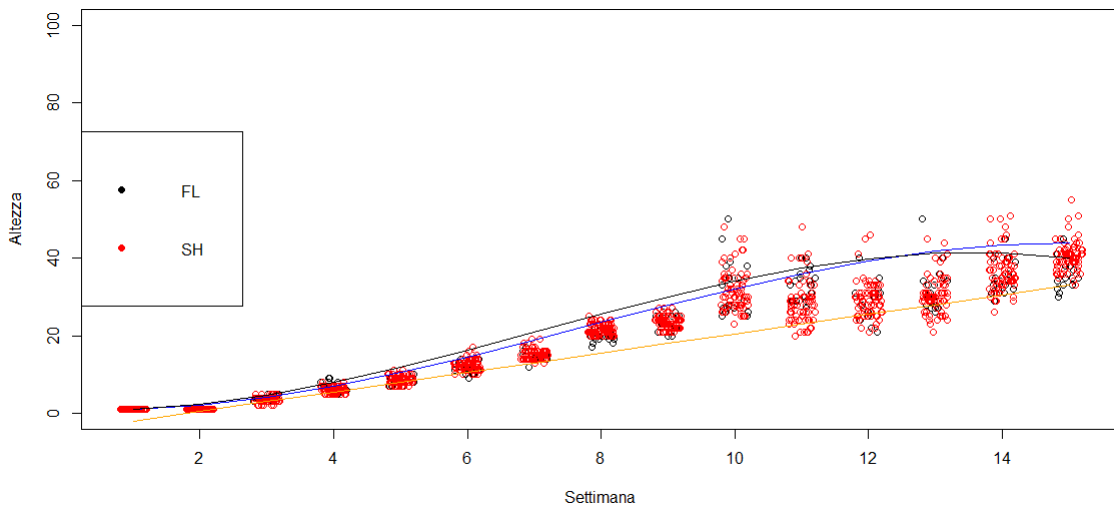


Figura 3.9. Curve di regressione quantilica per il quantile 0.10 (in arancione) e per quantile 0.90 (in blu per SH e in nero per FL)

I modelli con variabile risposta il numero di foglie hanno prodotti i seguenti risultati:

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d1 + d1:tempo +
d1:tempo2 + d1:tempo3, tau = taus, data = erba.medica)
```

tau: [1] 0.1

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-1.13333	0.42198	-2.68575	0.00733
tempo	0.31111	0.30857	1.00825	0.31352
tempo2	0.42222	0.06051	6.97779	0.00000
tempo3	-0.02222	0.00314	-7.08545	0.00000
d1	-0.04848	1.83282	-0.02645	0.97890
tempo:d1	0.04646	1.52957	0.03038	0.97577
tempo2:d1	-0.01313	0.32414	-0.04051	0.96769
tempo3:d1	0.00101	0.01743	0.05795	0.95380

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d1 + d1:tempo +
d1:tempo2 + d1:tempo3, tau = taus, data = erba.medica)
```

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.62915	0.32246	5.05232	0.00000
tempo	-1.45305	0.24483	-5.93495	0.00000
tempo2	0.86556	0.04381	19.75644	0.00000
tempo3	-0.04165	0.00192	-21.66100	0.00000
d1	1.14942	0.71395	1.60994	0.10765
tempo:d1	-1.46909	0.55462	-2.64881	0.00817
tempo2:d1	0.33587	0.10545	3.18523	0.00148
tempo3:d1	-0.01620	0.00480	-3.37750	0.00075

Per il quantile di ordine 0.10 risultano significativi i coefficienti relativi all'intercetta, al tempo² e al tempo³.

Per il quantile di ordine 0.90 risultano significativi i coefficienti dell'intercetta, delle variabili temporali e dei termini di interazione. Se per ciascun modello vengono tolte le variabili non significative si ottengono i seguenti risultati:

```
Call: rq(formula = nfoglie ~ tempo2 + tempo3, tau = 0.1, data = erba.medica)
```

tau: [1] 0.1

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-0.67781	0.12974	-5.22455	0.00000
tempo2	0.46768	0.01634	28.61327	0.00000
tempo3	-0.02411	0.00130	-18.52298	0.00000

Call: `rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d1:tempo + d1:tempo2 + d1:tempo3, tau = 0.9, data = erba.medica)`

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	1.62915	0.28198	5.77763	0.00000
tempo	-1.45305	0.21935	-6.62427	0.00000
tempo2	0.86556	0.04002	21.62669	0.00000
tempo3	-0.04165	0.00177	-23.49092	0.00000
tempo:d1	-0.51467	0.21264	-2.42035	0.01564
tempo2:d1	0.16298	0.06085	2.67853	0.00748
tempo3:d1	-0.00858	0.00313	-2.74197	0.00619

Le curve di regressione quantilica vengono riportate nel grafico riportato nella Figura 3.10:

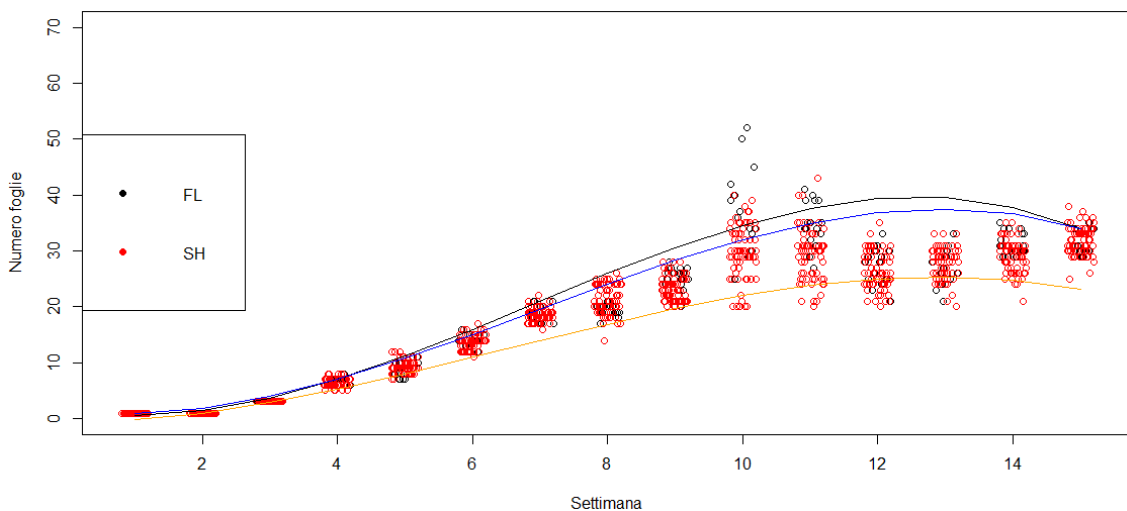


Figura 3.10. Curve di regressione quantilica per il quantile 0.10 (in arancione) e per il quantile 0.90 (in nero per FL e in blu per SH)

3.2 Coltura fagioli

3.2.1 Descrizione dei dati

Questo dataset riporta i valori di alcuni parametri di piante di fagiolo coltivate a diversi livelli di luce e ombra. Il dataset originale è una matrice 80×11 . Le colonne indicano rispettivamente:

- date: data in cui sono stati rilevati i valori dei parametri delle diverse piante. I parametri delle piante sono stati rilevati in un periodo compreso tra il 14 maggio 2021 e il 5 luglio 2021.
- week: settimana di riferimento. In totale sono state esaminate 5 settimane.
- light: condizione di luce o ombre in cui sono state coltivate le piante

- SH60: condizione di ombra al 60%
 - HSH60: condizione tra luce piena e ombra al 60%
 - SH30: condizione di ombra al 30%
 - FL (full light): in pieno sole.
- plant: numero della pianta. In ciascuna condizione (SH60, HSH60, SH30 e FL) sono state coltivate 4 piante.
 - height: altezza della pianta. Primo parametro rilevato.
 - diameter1: larghezza. Secondo parametro rilevato
 - diameter2: lunghezza. Terzo parametro rilevato.
 - number_leaves: numero delle foglie. Quarto parametro rilevato.
 - lenght_leaves: lunghezza delle foglie. Quinto parametro rilevato.
 - number_flowers: numero di fiori. Sesto parametro rilevato.
 - number_fruits: numero dei frutti. Settimo parametro rilevato.

Tuttavia, per semplificare le analisi i tre parametri height, diameter1, diameter2 sono stati uniti in un unico, calcolando il volume del fagiolo. Per fare questo la forma del fagiolo è stata approssimata a quella di un ellissoide e il volume è stato calcolato con la relativa formula:

$$V = \frac{4}{3} * \pi * \frac{height}{2} * \frac{diameter1}{2} * \frac{diameter2}{2} \quad (3.5)$$

I valori ottenuti sono stati arrotondati all'intero più vicino.

I grafici dell'andamento dei parametri "volume" e "number_leaves" nel corso delle settimane, per ciascuna delle quattro condizioni considerate (SH60, HSH60, SH30 e FL), vengono riportati di seguito. Quelli relativi alle restanti variabili vengono riportati in Appendice.

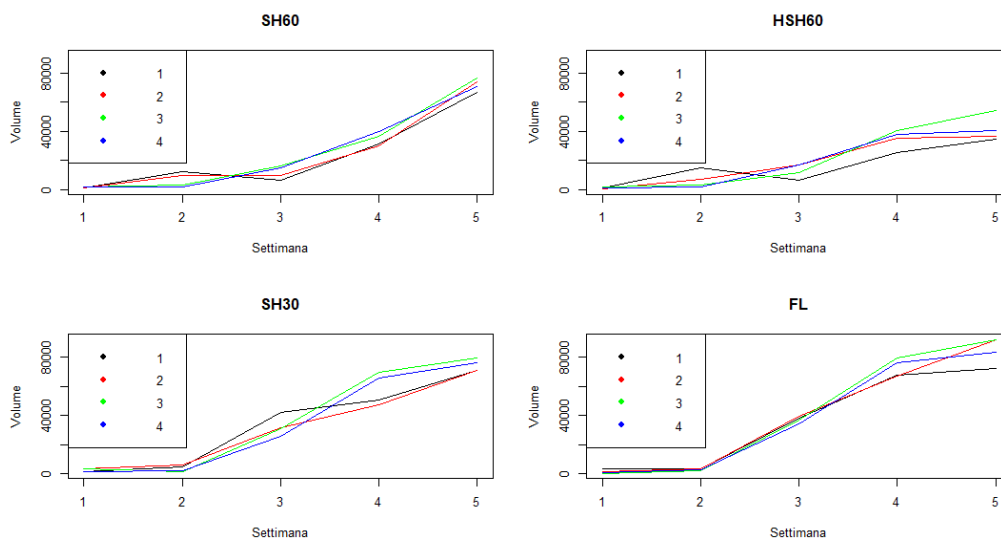


Figura 3.12. Grafici dell'andamento del numero delle foglie per i fagioli (SH60-HSH60-SH30-FL)

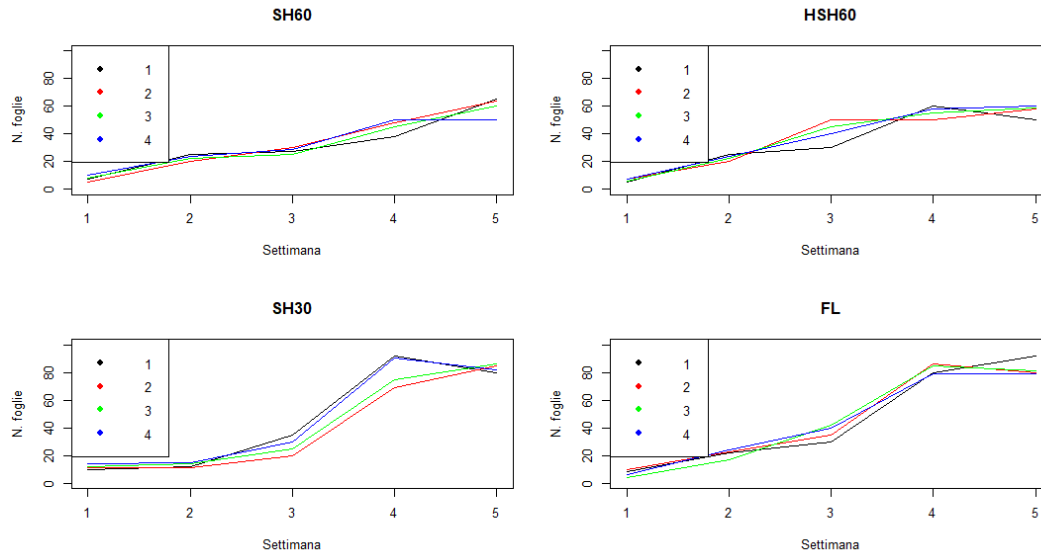


Figura 3.11. Grafici dell'andamento del volume per i fagioli (SH60-HSH60-SH30-FL)

3.2.2 Modelli di regressione con trend polinomiale

Nei prossimi modelli verranno considerate come variabili risposta “volume”, calcolato come definito precedentemente, “number_leaves”, “length_leaves”, “number_flowers” e “number_fruits”.

Come regressori, invece, si sono considerate le variabili “week” e “light”. Per inserire la variabile “light” nel modello sono state create quattro variabili dummy così definite:

$$x_{i1} = d1 = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{altrimenti} \end{cases} \quad (3.6) \quad x_{i2} = d2 = \begin{cases} 1 & \text{se light} = \text{SH60} \\ 0 & \text{altrimenti} \end{cases} \quad (3.7)$$

$$x_{i3} = d3 = \begin{cases} 1 & \text{se light} = \text{HSH60} \\ 0 & \text{altrimenti} \end{cases} \quad (3.8) \quad x_{i4} = d4 = \begin{cases} 1 & \text{se light} = \text{SH30} \\ 0 & \text{altrimenti} \end{cases} \quad (3.9)$$

Il numero di dummy da inserire come variabili indipendenti all'interno di un modello di regressione è sempre pari ad una in meno rispetto al numero di modalità della variabile qualitativa che si sta considerando. Se venissero inserite tutte, si otterrebbe la cosiddetta “trappola delle dummy”, causando problemi di multicollinearità.

La dummy esclusa dal modello viene detta di riferimento. Per tutti i modelli successivi si è scelta d1 come variabile di riferimento.

In primo luogo, si considera la variabile risposta volume.

Il modello con trend cubico ha prodotto i seguenti risultati:

```
Call:
lm(formula = volume ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-12719.0  -2485.1  -145.7   2561.8  12110.4
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	58536.0	13498.9	4.336	5.24e-05	***
d2	-64498.0	19090.3	-3.379	0.00124	**
d3	-41821.5	19090.3	-2.191	0.03212	*
d4	-18521.6	19090.3	-0.970	0.33559	
tempo	-96109.8	17645.1	-5.447	8.72e-07	***
tempo2	43573.3	6549.2	6.653	7.52e-09	***
tempo3	-4660.9	723.1	-6.446	1.73e-08	***
d2:tempo	107790.4	24953.9	4.320	5.55e-05	***
d2:tempo2	-49074.4	9262.0	-5.298	1.54e-06	***
d2:tempo3	5918.9	1022.6	5.788	2.33e-07	***
d3:tempo	69488.0	24953.9	2.785	0.00704	**
d3:tempo2	-30859.2	9262.0	-3.332	0.00144	**
d3:tempo3	3386.0	1022.6	3.311	0.00153	**
d4:tempo	31330.9	24953.9	1.256	0.21384	
d4:tempo2	-13739.7	9262.0	-1.483	0.14287	
d4:tempo3	1557.7	1022.6	1.523	0.13263	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5488 on 64 degrees of freedom
 Multiple R-squared: 0.9706, Adjusted R-squared: 0.9637
 F-statistic: 141 on 15 and 64 DF, p-value: < 2.2e-16

Il modello risulta: $\hat{y}_i = 58536.0 - 64498.0 \cdot d2 - 41821.5 \cdot d3 - 18521.6 \cdot d4 - 96109.8 \cdot \text{tempo} + 43573.3 \cdot \text{tempo}^2 - 4660.9 \cdot \text{tempo}^3 + 107790.4 \cdot d2 \cdot \text{tempo} - 49074.4 \cdot d2 \cdot \text{tempo}^2 + 5918.9 \cdot d2 \cdot \text{tempo}^3 + 69488.0 \cdot d3 \cdot \text{tempo} - 30859.2 \cdot d3 \cdot \text{tempo}^2 + 3386.0 \cdot d3 \cdot \text{tempo}^3 + 31330.9 \cdot d4 \cdot \text{tempo} - 13739.7 \cdot d4 \cdot \text{tempo}^2 + 1557.7 \cdot d4 \cdot \text{tempo}^3$

Considerando separatamente le diverse condizioni FL, SH60, HSH60, SH30

$\hat{y}_i = 58536.0 - 96109.8 \cdot \text{tempo} + 43573.3 \cdot \text{tempo}^2 - 4660.9 \cdot \text{tempo}^3$ per FL

$\hat{y}_i = (58536.0 - 64498.0) - (96109.8 + 107790.4) \cdot \text{tempo} + (43573.3 - 49074.4) \cdot \text{tempo}^2 - (4660.9 + 5918.9) \cdot \text{tempo}^3$ per SH60

$\hat{y}_i = (58536.0 - 41821.5) - (96109.8 + 69488.0) \cdot \text{tempo} + (43573.3 - 30859.2) \cdot \text{tempo}^2 - (4660.9 + 3386.0) \cdot \text{tempo}^3$ per HSH60

$\hat{y}_i = (58536.0 - 18521.6) - (96109.8 + 31330.9) \cdot \text{tempo} + (43573.3 - 13739.7) \cdot \text{tempo}^2 - (4660.9 + 1557.7) \cdot \text{tempo}^3$ per SH30

Dai risultati si nota che i termini di interazione con la variabile dummy d4 non risultano significativi; dunque, si accetta l'ipotesi di nullità di questi coefficienti. Inoltre, anche il coefficiente relativo alla variabile d4 non risulta significativo. I risultati indicano che non vi è una differenza significativa nella stima del trend tra la condizione FL e SH30.

Il coefficiente di determinazione R^2 indica che il modello spiega il 97.06% della variabilità della risposta.

Il grafico rappresentante le curve di regressione stimate dal modello viene riportato in Figura 3.13.

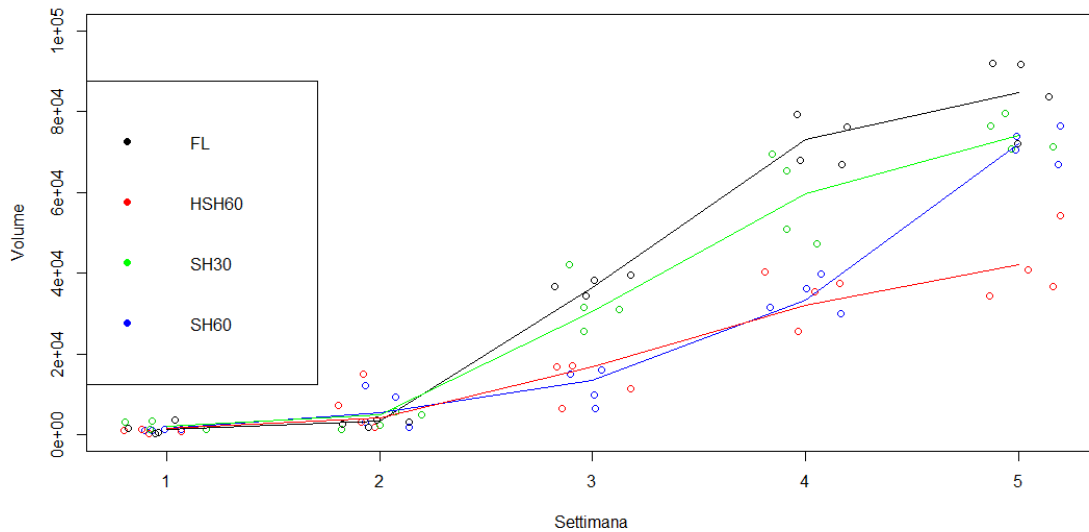


Figura 3.13. Diagramma di dispersione tra tempo e volume per i fagioli e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

In secondo luogo, si è considerata come variabile risposta il numero di foglie.

I risultati relativi al modello stimato sono:

```
Call:
lm(formula = nfoglie ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-17.7000  -3.4643   0.0018   3.4696  17.0500
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)    47.900     16.838   2.845 0.005963 **
d2             -58.050     23.813  -2.438 0.017566 *
d3             -47.350     23.813  -1.988 0.051044 .
d4              44.550     23.813   1.871 0.065940 .
tempo          -70.238     22.010  -3.191 0.002196 **
tempo2         34.679      8.169   4.245 7.21e-05 ***
tempo3         -3.833      0.902  -4.250 7.09e-05 ***
d2:tempo       92.482     31.127   2.971 0.004175 **
d2:tempo2     -39.330     11.553  -3.404 0.001150 **
d2:tempo3      4.438      1.276   3.479 0.000912 ***
d3:tempo       68.155     31.127   2.190 0.032208 *
d3:tempo2     -25.554     11.553  -2.212 0.030557 *
d3:tempo3      2.542      1.276   1.992 0.050585 .
d4:tempo      -55.637     31.127  -1.787 0.078604 .
d4:tempo2     17.696     11.553   1.532 0.130518
d4:tempo3     -1.667      1.276  -1.307 0.196040
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.846 on 64 degrees of freedom
Multiple R-squared: 0.9488, Adjusted R-squared: 0.9368
F-statistic: 79.08 on 15 and 64 DF, p-value: < 2.2e-16

Il modello risulta: $\hat{y}_i = 47.900 - 58.050 \cdot d2 - 47.350 \cdot d3 + 44.550 \cdot d4 - 70.238 \cdot \text{tempo} + 34.679 \cdot \text{tempo}^2 - 3.833 \cdot \text{tempo}^3 + 92.482 \cdot d2 \cdot \text{tempo} - 39.330 \cdot d2 \cdot \text{tempo}^2 + 4.438 \cdot d2 \cdot \text{tempo}^3 + 68.155 \cdot d3 \cdot \text{tempo} - 25.554 \cdot d3 \cdot \text{tempo}^2 + 2.542 \cdot d3 \cdot \text{tempo}^3 - 55.637 \cdot d4 \cdot \text{tempo} + 17.696 \cdot d4 \cdot \text{tempo}^2 - 1.667 \cdot d4 \cdot \text{tempo}^3$

Anche in questo caso i coefficienti dei termini di interazione tra d4 e tempo² e tra d4 e tempo³ non risultano significativi, mentre il termine di interazione tra d4 e tempo risulta debolmente significativo.

Il modello spiega il 94.88% della variabilità della risposta.

La rappresentazione delle curve di regressione stimate dal modello è riportata in Figura 3.14.

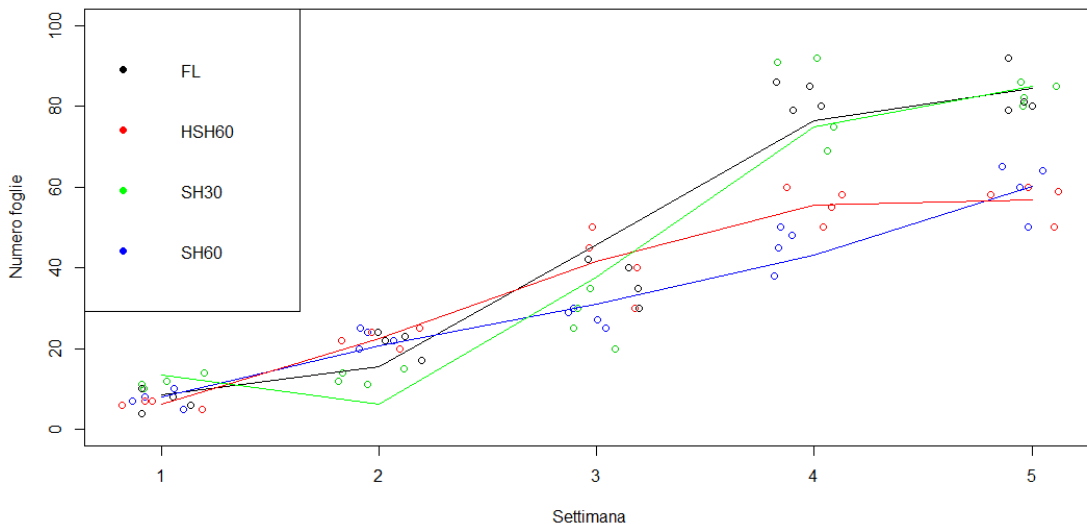


Figura 3.14. Diagramma di dispersione tra tempo e numero di foglie per i fagioli e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Dal grafico si può notare come le curve di regressione stimate per FL e SH30 abbiano un andamento molto simile, e nelle ultime settimane siano quasi coincidenti.

Successivamente si è considerata come variabile risposta la lunghezza delle foglie.

I risultati prodotti dal modello sono

```
Call:
lm(formula = lfoglie ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.63571 -0.68571  0.07857  0.61429  2.86429
```

```
Coefficients:
(Intercept)  1.40000    2.60376    0.538    0.5927
d2           -3.05000    3.68228   -0.828    0.4106
d3            1.45000    3.68228    0.394    0.6951
d4            0.95000    3.68228    0.258    0.7972
tempo         3.22619    3.40351    0.948    0.3467
tempo2       -0.48214    1.26327   -0.382    0.7040
tempo3        0.04167    0.13948    0.299    0.7661
d2:tempo      7.86310    4.81329    1.634    0.1072
d2:tempo2    -3.05357    1.78653   -1.709    0.0923 .
d2:tempo3     0.33333    0.19725    1.690    0.0959 .
d3:tempo      2.27976    4.81329    0.474    0.6374
```

```

d3:tempo2  -1.30357    1.78653   -0.730    0.4683
d3:tempo3   0.16667    0.19725    0.845    0.4013
d4:tempo    0.11310    4.81329    0.023    0.9813
d4:tempo2   0.50893    1.78653    0.285    0.7767
d4:tempo3  -0.10417    0.19725   -0.528    0.5993

```

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.059 on 64 degrees of freedom
 Multiple R-squared: 0.8505, Adjusted R-squared: 0.8155
 F-statistic: 24.28 on 15 and 64 DF, p-value: < 2.2e-16

Il modello risulta: $\hat{y}_i = 1.40 - 3.050 \cdot d2 + 1.45 \cdot d3 + 0.95 \cdot d4 + 3.22619 \cdot \text{tempo} + 0.48214 \cdot \text{tempo}^2 + 0.04167 \cdot \text{tempo}^3 + 7.86310 \cdot d2 \cdot \text{tempo} - 3.05357 \cdot d2 \cdot \text{tempo}^2 + 0.33333 \cdot d2 \cdot \text{tempo}^3 + 2.27976 \cdot d3 \cdot \text{tempo} - 1.30357 \cdot d3 \cdot \text{tempo}^2 + 0.16667 \cdot d3 \cdot \text{tempo}^3 + 0.11310 \cdot d4 \cdot \text{tempo} + 0.50893 \cdot d4 \cdot \text{tempo}^2 - 0.10417 \cdot d4 \cdot \text{tempo}^3$

In questo caso nessuna variabile risulta significativa, tranne i termini di interazione $d2 \times \text{tempo}^2$ e $d2 \times \text{tempo}^3$ i quali risultano debolmente significativi.

Il coefficiente R^2 indica che il modello ha un buon adattamento.

Il grafico delle curve stimate dal modello è riportato in Figura 3.15.

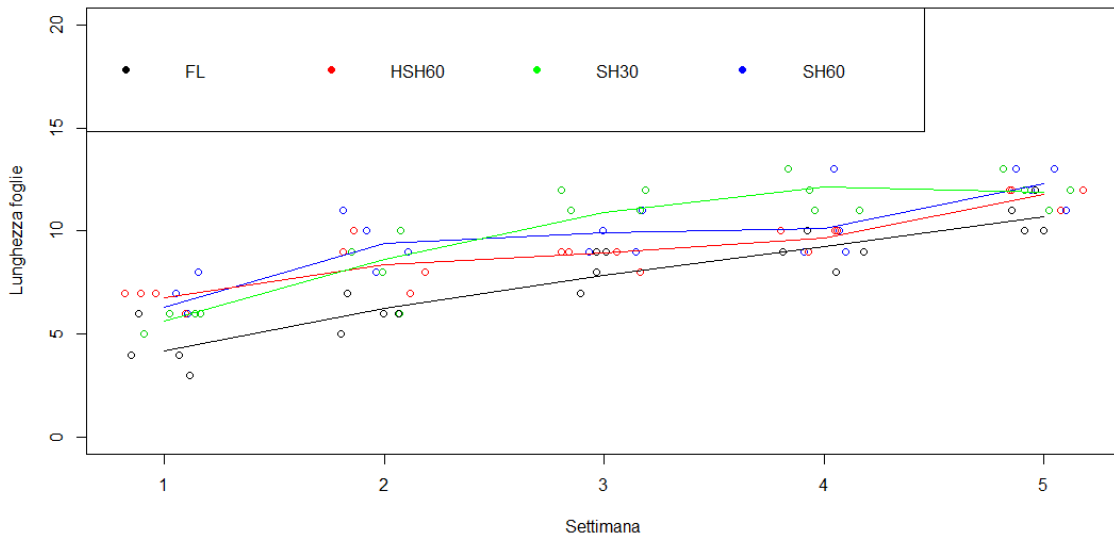


Figura 3.15. Diagramma di dispersione tra tempo e lunghezza delle foglie per i fagioli e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Dopodiché si è considerata come variabile dipendente il numero di fiori.

Il modello ha fornito i seguenti risultati:

```

Call:
lm(formula = nfiori ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)

```

Residuals:

Min	1Q	Median	3Q	Max
-4.1857	-1.0875	0.3643	0.4714	4.0286

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.65000	5.15697	-1.096	0.277
d2	-4.45000	7.29306	-0.610	0.544
d3	-2.75000	7.29306	-0.377	0.707
d4	3.65000	7.29306	0.500	0.618
tempo	6.00595	6.74094	0.891	0.376
tempo2	-0.78571	2.50201	-0.314	0.755
tempo3	-0.04167	0.27625	-0.151	0.881
d2:tempo	6.20833	9.53313	0.651	0.517
d2:tempo2	-1.87500	3.53837	-0.530	0.598
d2:tempo3	0.16667	0.39068	0.427	0.671
d3:tempo	3.63690	9.53313	0.382	0.704
d3:tempo2	-0.82143	3.53837	-0.232	0.817
d3:tempo3	0.04167	0.39068	0.107	0.915
d4:tempo	-5.95238	9.53313	-0.624	0.535
d4:tempo2	2.71429	3.53837	0.767	0.446
d4:tempo3	-0.33333	0.39068	-0.853	0.397

Residual standard error: 2.097 on 64 degrees of freedom
Multiple R-squared: 0.6374, Adjusted R-squared: 0.5525
F-statistic: 7.501 on 15 and 64 DF, p-value: 3.161e-0

Anche per questo modello tutte le variabili risultano non significative.

Il grafico con le curve di regressione viene riportato sotto in Figura 3.16.

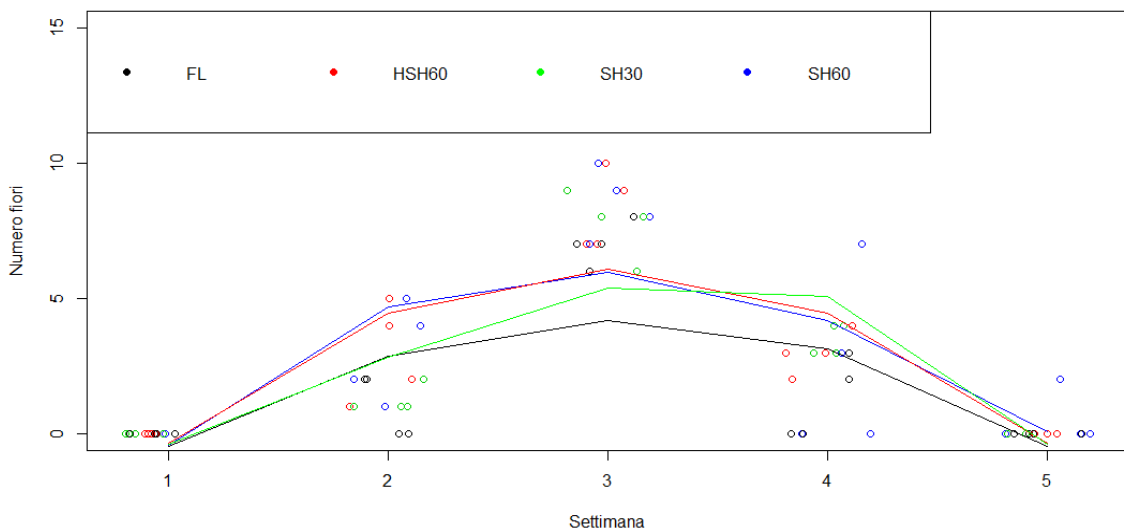


Figura 3.16. Diagramma di dispersione tra tempo e numero di fiori per i fagioli e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Infine, si è considerata come variabile dipendente nel modello il numero di frutti, i cui risultati sono:

```
Call:
lm(formula = nfrutti ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.4714	-1.2321	-0.2643	1.3643	7.5643

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.40000	5.37294	2.308	0.0243 *
d2	-2.75000	7.59849	-0.362	0.7186
d3	-6.20000	7.59849	-0.816	0.4176
d4	2.45000	7.59849	0.322	0.7482

tempo	-17.54762	7.02325	-2.499	0.0151 *
tempo2	6.03571	2.60679	2.315	0.0238 *
tempo3	-0.41667	0.28782	-1.448	0.1526
d2: tempo	3.47024	9.93237	0.349	0.7279
d2: tempo2	-0.82143	3.68655	-0.223	0.8244
d2: tempo3	-0.04167	0.40704	-0.102	0.9188
d3: tempo	8.38095	9.93237	0.844	0.4019
d3: tempo2	-2.53571	3.68655	-0.688	0.4940
d3: tempo3	0.08333	0.40704	0.205	0.8384
d4: tempo	-3.86310	9.93237	-0.389	0.6986
d4: tempo2	1.67857	3.68655	0.455	0.6504
d4: tempo3	-0.20833	0.40704	-0.512	0.6105

 Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.184 on 64 degrees of freedom
 Multiple R-squared: 0.9354, Adjusted R-squared: 0.9203
 F-statistic: 61.82 on 15 and 64 DF, p-value: < 2.2e-16

In quest’ultimo caso risultano significativi solamente l’intercetta e i coefficienti delle potenze del tempo, suggerendo che la stima del trend è di tipo parabolico. Ciò è evidenziato anche dal grafico delle curve di regressione stimate dal modello, riportato in Figura 3.17.

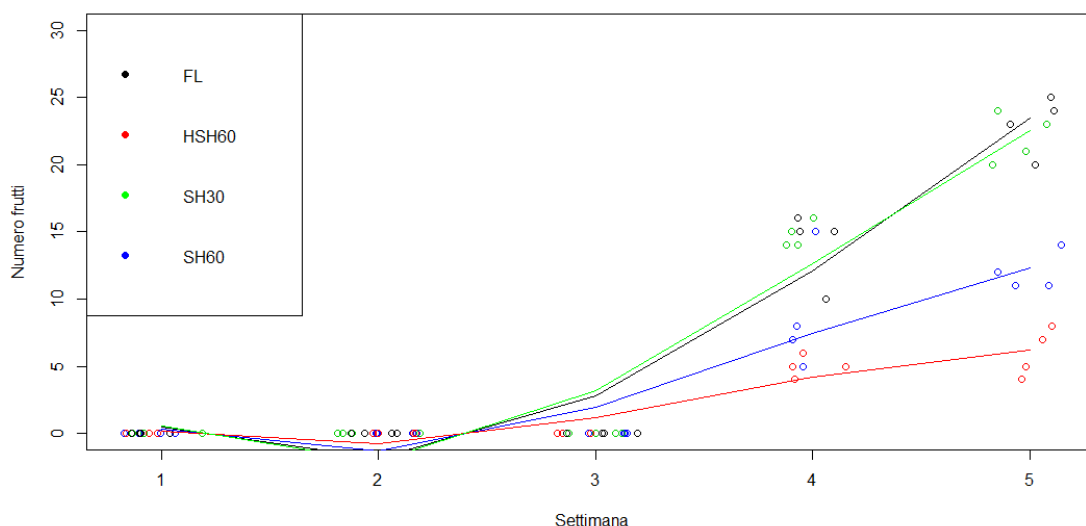


Figura 3.17. Diagramma di dispersione tra tempo e numero di frutti per i fagioli e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

3.2.3 Regressione quantilica

Nei seguenti modelli, analogamente ai precedenti, si prendono come variabili risposte “volume”, “number_leaves”, “length_leaves”, “number_flowers” e “number_fruits”.

Come regressori, invece, si sono considerate le variabili “week” e “light”. Per inserire la variabile “light” nel modello sono state create quattro variabili dummy così definite, con la stessa formulazione delle Formule 3.6, 3.7, 3.8, 3.9.

$$x_{i1} = d1 = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i2} = d2 = \begin{cases} 1 & \text{se light} = \text{SH60} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i3} = d3 = \begin{cases} 1 & \text{se light} = \text{HSH60} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i4} = d4 = \begin{cases} 1 & \text{se light} = \text{SH30} \\ 0 & \text{altrimenti} \end{cases}$$

Verranno analizzati il quantile di ordine 0.10 e 0.90, allo stesso modo dei precedenti modelli di regressione quantilica.

Il primo modello con variabile risposta il volume ha prodotto i seguenti risultati:

```
Call: rq(formula = volume ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.1
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	57931.66667	16017.55508	3.61676	0.00059
tempo	-96853.16667	22214.14526	-4.35998	0.00005
tempo2	44092.33333	8853.75811	4.98007	0.00001
tempo3	-4831.83333	1054.93886	-4.58020	0.00002
d2	-55192.00000	22783.34675	-2.42247	0.01826
d3	-13113.16667	35697.73507	-0.36734	0.71458
d4	-26559.00000	23169.93391	-1.14627	0.25595
tempo:d2	95344.50000	31501.23497	3.02669	0.00356
tempo2:d2	-45157.50000	12236.38980	-3.69043	0.00046
tempo3:d2	5619.00000	1412.99677	3.97665	0.00018
tempo:d3	28665.87500	52250.75362	0.54862	0.58517
tempo2:d3	-17714.33333	19986.34049	-0.88632	0.37876
tempo3:d3	2199.62500	2225.84320	0.98822	0.32677
tempo:d4	48053.33333	32296.68253	1.48787	0.14170
tempo2:d4	-23474.00000	12877.14559	-1.82292	0.07299
tempo3:d4	2974.66667	1512.46462	1.96677	0.05355

```
Call: rq(formula = volume ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.9
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	69065.50000	15473.24541	4.46354	0.00003
tempo	-108824.54167	21009.14329	-5.17987	0.00000
tempo2	48479.00000	8130.14604	5.96287	0.00000
tempo3	-5159.95833	925.43617	-5.57570	0.00000
d2	-89232.83333	23012.35722	-3.87761	0.00025
d3	-83781.16667	27069.89401	-3.09499	0.00292
d4	-14069.16667	22660.74833	-0.62086	0.53690
tempo:d2	139160.37500	31594.50381	4.40458	0.00004
tempo2:d2	-58857.91667	12035.80325	-4.89024	0.00001
tempo3:d2	6794.37500	1351.41859	5.02759	0.00000
tempo:d3	126616.70833	37929.00589	3.33826	0.00141
tempo2:d3	-50356.83333	14373.12331	-3.50354	0.00084
tempo3:d3	5375.29167	1630.42931	3.29686	0.00160
tempo:d4	21643.20833	31268.61055	0.69217	0.49133
tempo2:d4	-8662.33333	12236.83383	-0.70789	0.48158
tempo3:d4	879.29167	1398.84722	0.62858	0.53186

Per il quantile di ordine 0.10 risultano significativi i coefficienti relativi all'intercetta, alle variabili temporali, alla dummy d2 e ai termini di interazione ad essa associati. Non risultano significativi invece i termini di interazione relativi a d3 e d4. Dunque, non vi è una differenza significativa tra FL, HSH60 e SH30, mentre vi è per la condizione SH60.

Per il quantile di ordine 0.90 risultano significativi i coefficienti relativi all'intercetta, alle variabili temporali, alle dummy d2 e d3 e ai termini di interazione associati ad esse. Quindi le condizioni FL e SH30 non presentano differenze significative. Il relativo grafico è riportato in Figura 3.18.

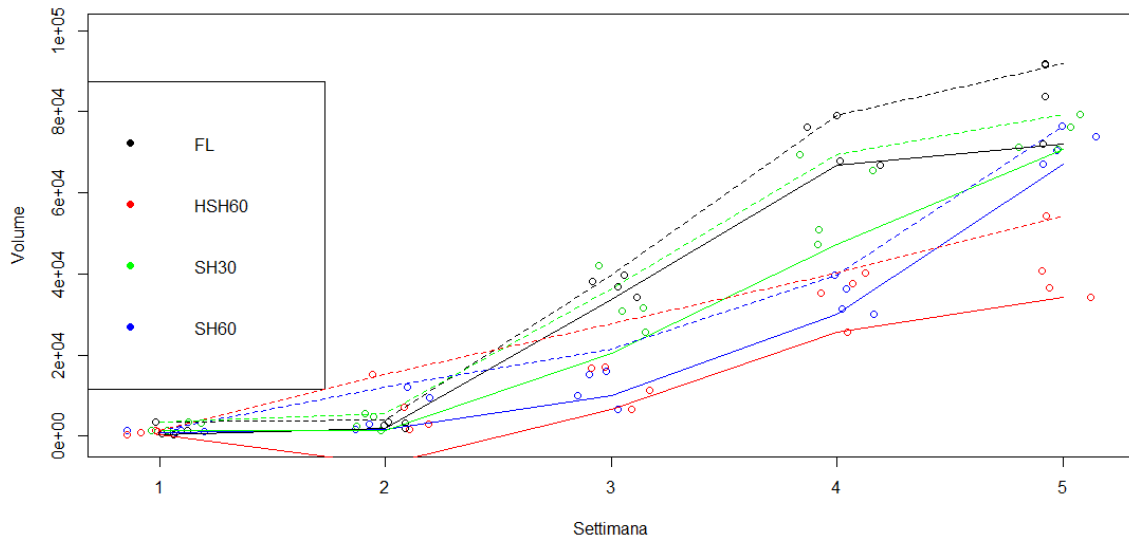


Figura 3.18. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta volume

Il secondo modello con variabile risposta il numero di foglie ha prodotto i seguenti risultati:

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

tau: [1] 0.1

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-14.75000	47.35176	-0.31150	0.75643
tempo	23.54167	72.65272	0.32403	0.74697
tempo2	-5.75000	30.52145	-0.18839	0.85117
tempo3	0.95833	3.57514	0.26805	0.78952
d2	17.25000	52.74155	0.32707	0.74468
d3	52.25000	58.08869	0.89949	0.37177
d4	39.75000	61.37415	0.64767	0.51952
tempo:d2	-24.66667	79.82723	-0.30900	0.75832
tempo2:d2	9.75000	33.04733	0.29503	0.76892
tempo3:d2	-1.33333	3.84773	-0.34652	0.73008
tempo:d3	-80.41667	87.50120	-0.91904	0.36153
tempo2:d3	33.25000	35.31459	0.94154	0.34997
tempo3:d3	-4.08333	4.04354	-1.00984	0.31637
tempo:d4	-47.54167	93.81418	-0.50676	0.61406
tempo2:d4	15.25000	39.32966	0.38775	0.69949
tempo3:d4	-1.45833	4.60234	-0.31687	0.75238

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	35.33333	17.29180	2.04336	0.04514
tempo	-52.00000	22.63003	-2.29783	0.02485
tempo2	30.16667	8.48353	3.55591	0.00071
tempo3	-3.50000	0.97684	-3.58299	0.00066

d2	-45.33333	23.82949	-1.90240	0.06162
d3	-24.08333	25.59909	-0.94079	0.35035
d4	57.33333	24.66245	2.32472	0.02327
tempo:d2	75.33333	31.35279	2.40276	0.01918
tempo2:d2	-33.91667	11.78530	-2.87788	0.00544
tempo3:d2	3.91667	1.34213	2.91825	0.00485
tempo:d3	31.54167	35.05446	0.89979	0.37161
tempo2:d3	-11.41667	13.66376	-0.83554	0.40652
tempo3:d3	0.95833	1.57482	0.60853	0.54499
tempo:d4	-80.16667	33.13443	-2.41944	0.01840
tempo2:d4	30.16667	12.94117	2.33106	0.02291
tempo3:d4	-3.33333	1.51628	-2.19836	0.03155

Per il quantile 0.10 nessuna variabile risulta statisticamente significativa.

Per il quantile 0.90 non risultano significative la dummy d3 e i relativi termini di interazione con il tempo.

Le curve di regressione quantilica stimate per i diversi quantili sono riportate di seguito in Figura 3.19

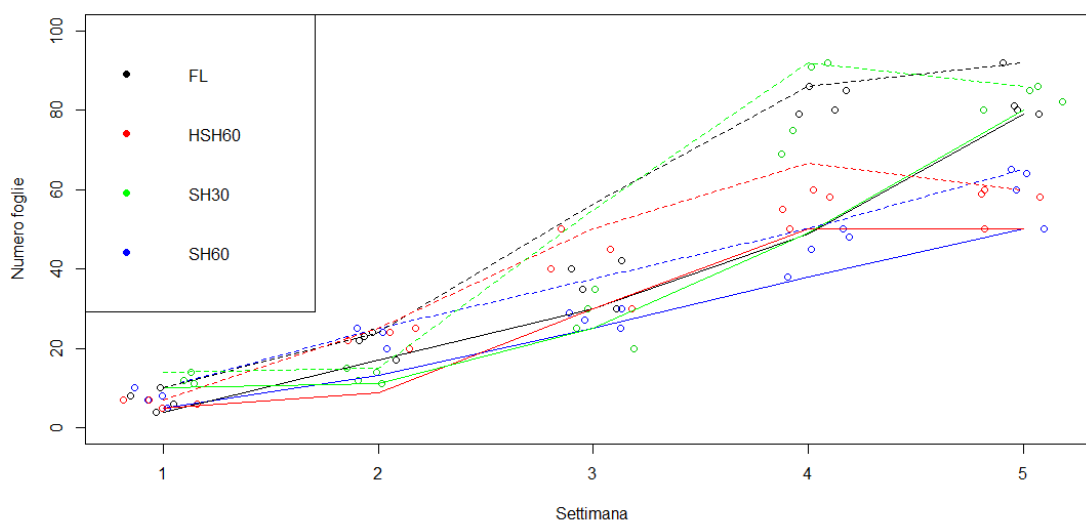


Figura 3.19. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta numero di foglie

I modelli con variabile dipendente “length_leaves” hanno prodotto i seguenti risultati:

```
Call: rq(formula = lfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

tau: [1] 0.1

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.00000	3.56481	0.00000	1.00000
tempo	3.66667	4.55275	0.80537	0.42359
tempo2	-0.75000	1.66139	-0.45143	0.65321
tempo3	0.08333	0.18180	0.45839	0.64823
d2	-5.66667	6.41378	-0.88351	0.38026
d3	4.33333	4.94852	0.87568	0.38448
d4	7.66667	6.40440	1.19709	0.23569
tempo:d2	9.66667	7.68028	1.25863	0.21273
tempo2:d2	-3.33333	2.72227	-1.22447	0.22526
tempo3:d2	0.33333	0.29287	1.13817	0.25929
tempo:d3	-1.50000	6.53903	-0.22939	0.81930
tempo2:d3	0.16667	2.40175	0.06939	0.94489

```

tempo3:d3    0.00000  0.26207   0.00000  1.00000
tempo:d4    -8.83333  8.70361  -1.01490  0.31397
tempo2:d4    3.58333  3.07609   1.16490  0.24838
tempo3:d4   -0.41667  0.32297  -1.29011  0.20166

```

```

Call: rq(formula = lfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)

```

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	2.00000	5.20465	0.38427	0.70205
tempo	5.33333	6.66414	0.80030	0.42650
tempo2	-1.50000	2.40994	-0.62242	0.53588
tempo3	0.16667	0.25972	0.64173	0.52334
d2	1.00000	7.56330	0.13222	0.89523
d3	-3.33333	6.22380	-0.53558	0.59411
d4	-2.33333	6.42170	-0.36335	0.71754
tempo:d2	0.83333	10.08298	0.08265	0.93439
tempo2:d2	0.25000	3.84560	0.06501	0.94837
tempo3:d2	-0.08333	0.43186	-0.19297	0.84760
tempo:d3	6.50000	8.15264	0.79729	0.42823
tempo2:d3	-2.41667	2.95665	-0.81737	0.41675
tempo3:d3	0.25000	0.31844	0.78509	0.43530
tempo:d4	2.33333	8.54325	0.27312	0.78564
tempo2:d4	0.08333	3.15290	0.02643	0.97900
tempo3:d4	-0.08333	0.34379	-0.24240	0.80925

In questo caso non risulta significativa nessuna variabile per entrambi i quantili.

Le curve di regressione quantilica stimate per i diversi quantili sono riportate di seguito in Figura 3.20

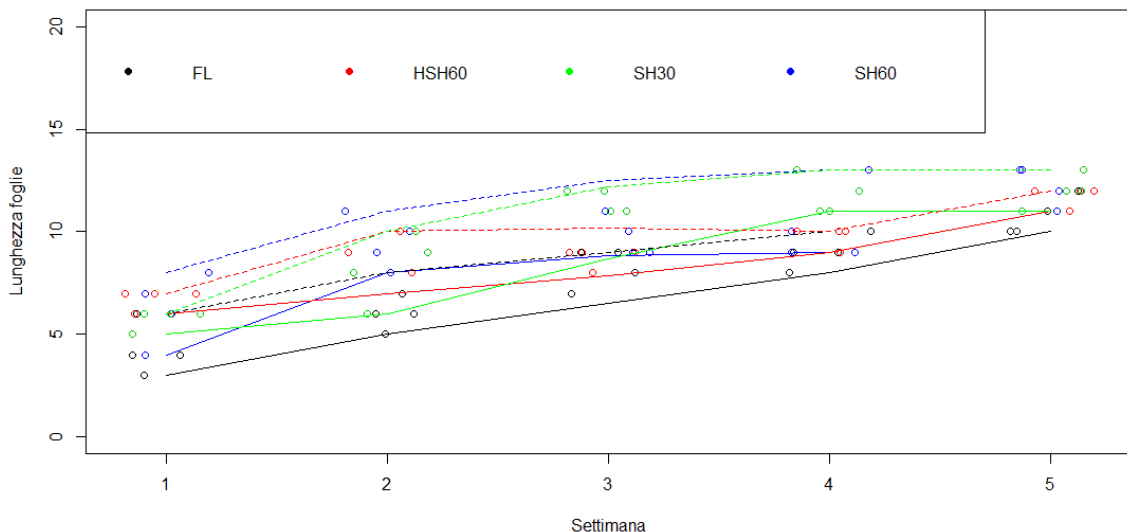


Figura 3.20. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta lunghezza delle foglie

I risultati dei modelli con variabile risposta il numero di fiori sono i seguenti

```

Call: rq(formula = nfiori ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)

```

tau: [1] 0.1

Coefficients:

	Value	Std. Error	t value	Pr(> t)
--	-------	------------	---------	----------

```

(Intercept)  0.00000  5.52598    0.00000  1.00000
tempo        0.00000  7.28883    0.00000  1.00000
tempo2       0.00000  2.66148    0.00000  1.00000
tempo3       0.00000  0.29414    0.00000  1.00000
d2          -3.33333  8.55610   -0.38959  0.69814
d3           0.00000  8.37325    0.00000  1.00000
d4           1.66667  7.47016    0.22311  0.82416
tempo:d2     4.83333 11.52784    0.41927  0.67642
tempo2:d2   -1.66667  4.24077   -0.39301  0.69562
tempo3:d2    0.16667  0.46982    0.35475  0.72395
tempo:d3    -0.83333 11.15602   -0.07470  0.94069
tempo2:d3    1.00000  4.00412    0.24974  0.80359
tempo3:d3   -0.16667  0.43322   -0.38471  0.70172
tempo:d4    -3.66667  9.77052   -0.37528  0.70869
tempo2:d4    2.33333  3.56011    0.65541  0.51455
tempo3:d4   -0.33333  0.39254   -0.84917  0.39895

```

```

Call: rq(formula = nfiori ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)

```

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	7.50000	9.04334	0.82934	0.40999
tempo	-14.41667	13.75706	-1.04795	0.29860
tempo2	8.00000	5.65956	1.41354	0.16234
tempo3	-1.08333	0.65363	-1.65741	0.10233
d2	-10.50000	12.01315	-0.87404	0.38536
d3	-10.00000	14.68956	-0.68076	0.49848
d4	2.50000	12.41916	0.20130	0.84110
tempo:d2	15.41667	18.22106	0.84609	0.40065
tempo2:d2	-5.50000	7.43346	-0.73990	0.46207
tempo3:d2	0.58333	0.85440	0.68274	0.49723
tempo:d3	14.50000	22.50047	0.64443	0.52160
tempo2:d3	-5.00000	9.25541	-0.54022	0.59092
tempo3:d3	0.50000	1.06829	0.46804	0.64135
tempo:d4	-4.25000	18.91464	-0.22469	0.82293
tempo2:d4	2.00000	7.82499	0.25559	0.79909
tempo3:d4	-0.25000	0.90732	-0.27554	0.78379

Anche per il numero di fiori non risulta significativa nessuna variabile per entrambi i quantili analizzati. Il relativo grafico con le curve di regressione stimate è rappresentato in Figura 3.21.

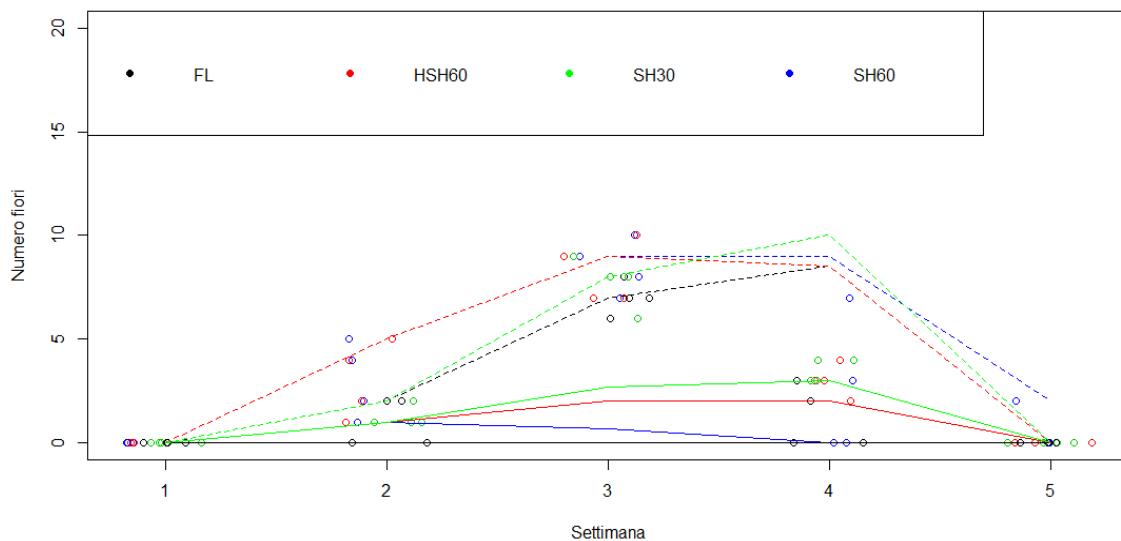


Figura 3.21. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta numero di fiori

Infine, come ultima variabile risposta si è considerato il numero di frutti della pianta.

I modelli stimati per i due quantili considerati hanno prodotto i seguenti risultati:

```
Call: rq(formula = nfrutti ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.1
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	20.00000	15.44672	1.29477	0.20005
tempo	-29.16667	23.35452	-1.24887	0.21626
tempo2	10.00000	8.99588	1.11162	0.27046
tempo3	-0.83333	0.99337	-0.83889	0.40465
d2	-11.50000	16.71513	-0.68800	0.49394
d3	-6.00000	16.71906	-0.35887	0.72087
d4	-25.00000	19.12105	-1.30746	0.19573
tempo:d2	16.95833	25.04092	0.67722	0.50070
tempo2:d2	-6.00000	9.64384	-0.62216	0.53605
tempo3:d2	0.54167	1.06649	0.50790	0.61327
tempo:d3	8.00000	25.03957	0.31949	0.75039
tempo2:d3	-2.00000	9.61207	-0.20807	0.83583
tempo3:d3	0.00000	1.06093	0.00000	1.00000
tempo:d4	38.33333	28.87959	1.32735	0.18911
tempo2:d4	-15.00000	11.41099	-1.31452	0.19336
tempo3:d4	1.66667	1.28417	1.29785	0.19900

```
Call: rq(formula = nfrutti ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.9
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	10.00000	5.22960	1.91219	0.06033
tempo	-16.16667	6.95132	-2.32570	0.02322
tempo2	6.75000	2.66726	2.53069	0.01385
tempo3	-0.58333	0.30885	-1.88871	0.06346
d2	5.66667	9.47159	0.59828	0.55176
d3	-5.33333	7.20964	-0.73975	0.46216
d4	0.66667	7.34069	0.09082	0.92792
tempo:d2	-10.00000	14.32770	-0.69795	0.48774
tempo2:d2	5.08333	6.33653	0.80223	0.42539
tempo3:d2	-0.75000	0.78262	-0.95832	0.34151
tempo:d3	8.50000	9.55919	0.88920	0.37723
tempo2:d3	-3.41667	3.63651	-0.93954	0.35098
tempo3:d3	0.25000	0.41624	0.60062	0.55021
tempo:d4	-1.16667	9.68035	-0.12052	0.90445
tempo2:d4	0.58333	3.66759	0.15905	0.87413
tempo3:d4	-0.08333	0.42164	-0.19764	0.84395

Per il quantile di ordine 0.10 non risulta significativa nessuna variabile.

Per il quantile di ordine 0.90, invece, risultano significativi i coefficienti delle variabili tempo e tempo². Il grafico con le curve di regressione quantilica stimate viene riportato in Figura 3.22.

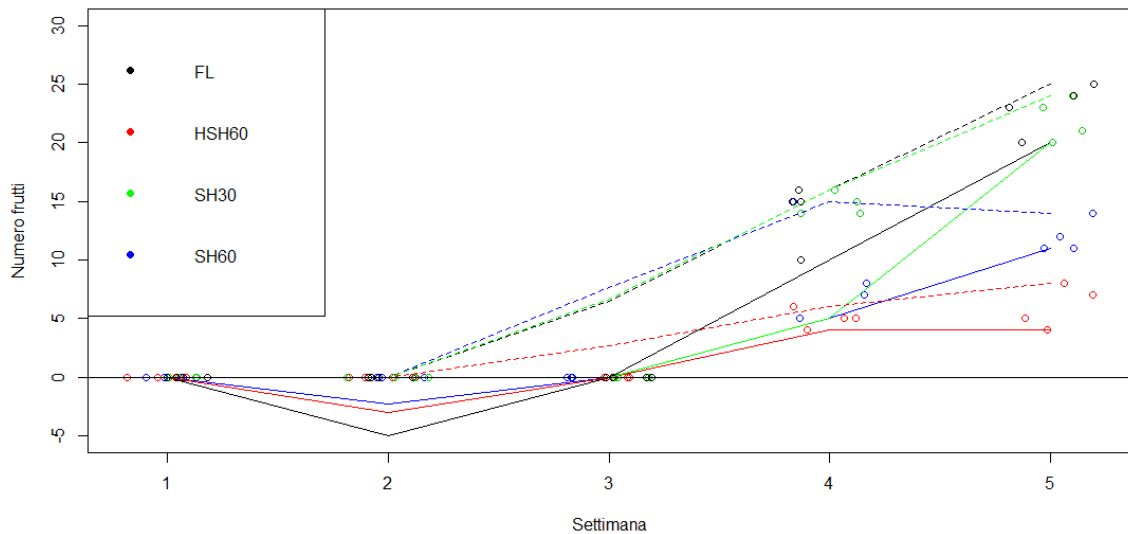


Figura 3.22. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta numero di frutti

3.3 Coltura patate

3.3.1 Descrizione dei dati

L'ultimo dataset riporta i valori di parametri della pianta di patate. Anche queste piante sono state coltivate in diverse condizioni di sole e ombreggiamento. Il dataset è costituito da 384 rilevazioni (righe) rispetto 11 variabili (colonne). Nelle colonne sono riportati:

- date: data in cui sono stati rilevati i valori dei parametri delle diverse piante. I parametri delle piante sono stati rilevati in un periodo compreso tra il 28 aprile 2021 e il 5 luglio 2021.
- week: settimana di riferimento. In totale sono state esaminate 6 settimane.
- light: condizione di luce o ombre in cui sono state coltivate le piante
 - FL (full light): in pieno sole
 - SH30: condizione di ombre al 30%
 - HSH60: condizione tra luce piena e ombra al 60%
 - SH60: condizione di ombra al 60%
- area: area di riferimento. Ciascuna condizione di luce e ombra è stata divisa in 4 parti, per un totale di 16 aree.
- plant: numero della pianta. In ciascuna area sono state coltivate 4 piante.
- height: altezza della pianta. Primo parametro rilevato.
- diameter1: larghezza. Secondo parametro rilevato

- diameter2: lunghezza. Terzo parametro rilevato.
- number_leaves: numero delle foglie. Quarto parametro rilevato.
- lenght_leaves: lunghezza delle foglie. Quinto parametro rilevato.
- number_flowers: numero di fiori. Sesto parametro rilevato.

Anche per le patate, per semplificare l'analisi, i tre parametri height, diameter1, diameter2, sono stati raggruppati calcolando il volume della patata, la cui forma è stata approssimata a quella di un ellissoide, secondo la formula precedente

$$V = \frac{4}{3} * \pi * \frac{height}{2} * \frac{diameter1}{2} * \frac{diameter2}{2} \quad (3.10)$$

I grafici dell'andamento del volume nel corso delle settimane, per ciascuna area, vengono riportati di seguito. Invece, i grafici relativi alle altre tre variabili, numero delle foglie, lunghezza delle foglie e numero dei fiori vengono riportati in Appendice.

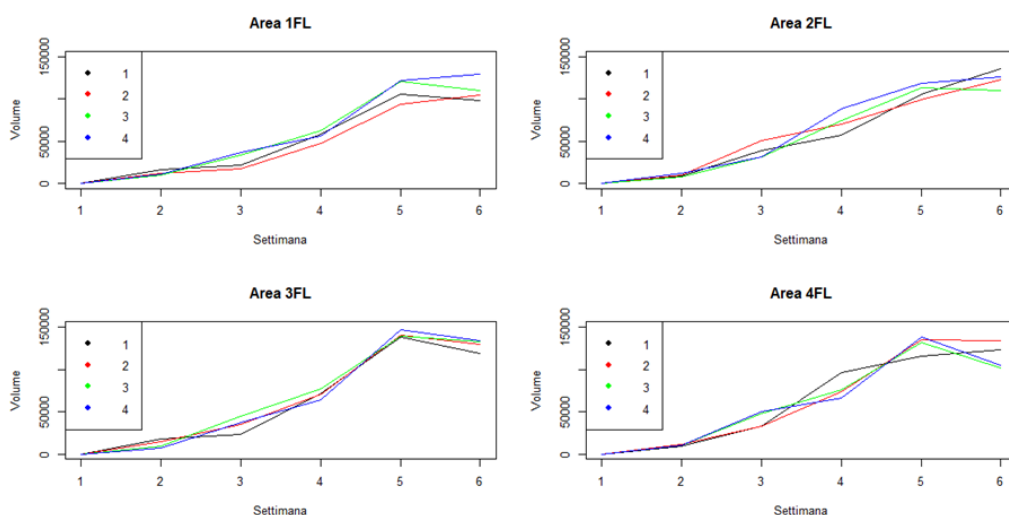


Figura 3.23. Grafici dell'andamento del volume per le patate per aree FL

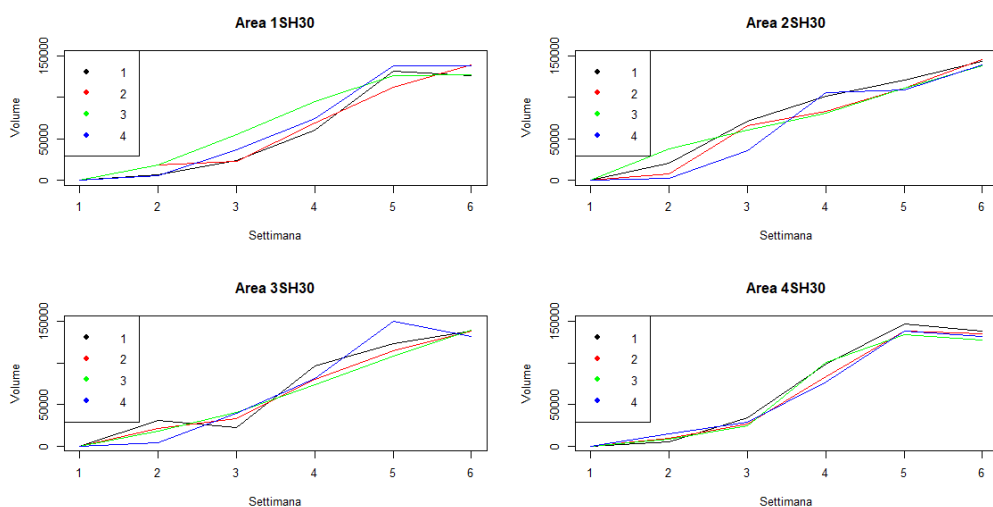


Figura 3.24. Grafici dell'andamento del volume per le patate per aree SH30

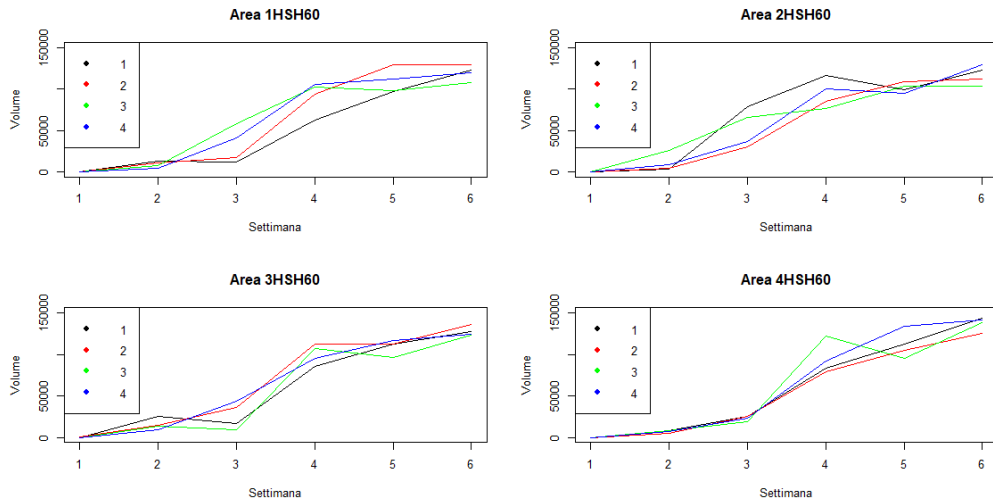


Figura 3.25. Grafici dell'andamento del volume per le patate per aree HSH60

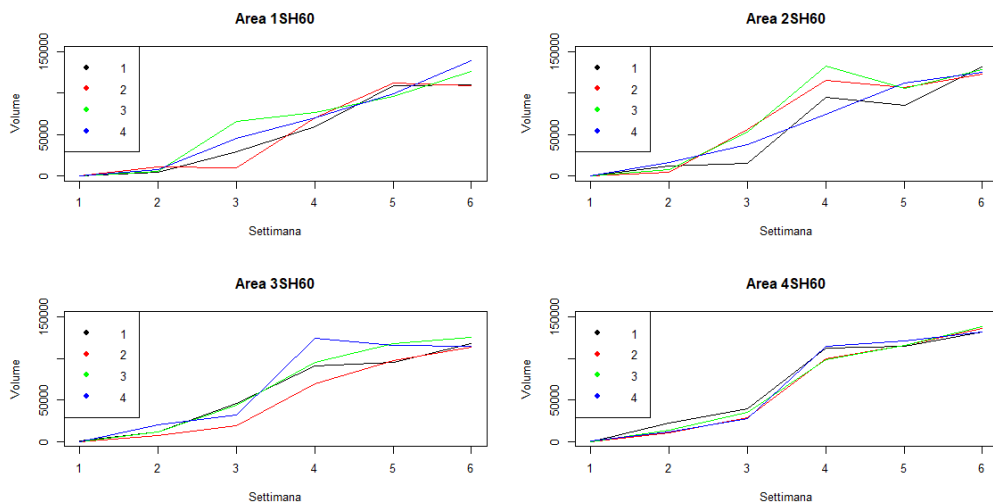


Figura 3.26. Grafici dell'andamento del volume per le patate per aree SH60

3.3.2 Modelli di regressione con trend polinomiale

Nei successivi modelli verranno considerate come variabili risposte “volume”, calcolato come definito precedentemente, “number_leaves”, “length_leaves” e “number_flowers”.

Come regressori, invece, si sono considerate le variabili “week” e “light”. Per inserire la variabile “light” nel modello sono state create quattro variabili dummy così definite, allo stesso modo delle Formule 3.6, 3.7, 3.8,3.9:

$$x_{i1} = d1 = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i2} = d2 = \begin{cases} 1 & \text{se light} = \text{SH60} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i3} = d3 = \begin{cases} 1 & \text{se light} = \text{HSH60} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i4} = d4 = \begin{cases} 1 & \text{se light} = \text{SH30} \\ 0 & \text{altrimenti} \end{cases}$$

Come per i modelli precedenti, si è considerata come variabile di riferimento d1.

Per prima cosa si considera come variabile risposta il volume.

I risultati del modello stimato sono:

```
Call:
lm(formula = volume ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-33389  -6039    -90     6168   49561
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  45291.7    11786.6   3.843  0.000143 ***
d2           -17121.8    16668.7  -1.027  0.305010
d3           -11949.5    16668.7  -0.717  0.473903
d4           -10061.6    16668.7  -0.604  0.546467
tempo        -71998.4    13429.7  -5.361  1.46e-07 ***
tempo2         31927.1     4297.4   7.429  7.69e-13 ***
tempo3        -2962.2     406.1   -7.294  1.86e-12 ***
d2:tempo      19352.4    18992.5   1.019  0.308896
d2:tempo2    -4965.4     6077.5  -0.817  0.414454
d2:tempo3     371.5      574.3   0.647  0.518137
d3:tempo      12389.1    18992.5   0.652  0.514605
d3:tempo2    -2744.4     6077.5  -0.452  0.651845
d3:tempo3     172.1      574.3   0.300  0.764626
d4:tempo      12252.8    18992.5   0.645  0.519238
d4:tempo2    -3163.4     6077.5  -0.521  0.603024
d4:tempo3     300.5      574.3   0.523  0.601131
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
Residual standard error: 13080 on 368 degrees of freedom
Multiple R-squared: 0.9375, Adjusted R-squared: 0.9349
F-statistic: 367.9 on 15 and 368 DF, p-value: < 2.2e-16
```

Il modello risulta: $\hat{y}_i = 45291.7 - 17121.8 \cdot d2 - 11949.5 \cdot d3 - 10061.6 \cdot d4 - 71998.4 \cdot \text{tempo} + 31927.1 \cdot \text{tempo}^2 - 2962.2 \cdot \text{tempo}^3 + 19352.4 \cdot d2 \cdot \text{tempo} - 4965.4 \cdot d2 \cdot \text{tempo}^2 + 371.5 \cdot d2 \cdot \text{tempo}^3 + 12389.1 \cdot d3 \cdot \text{tempo} - 2744.4 \cdot d3 \cdot \text{tempo}^2 + 172.1 \cdot d3 \cdot \text{tempo}^3 + 12252.8 \cdot d4 \cdot \text{tempo} - 3163.4 \cdot d4 \cdot \text{tempo}^2 + 300.5 \cdot \text{tempo}^3$

Considerando separatamente ciascuna condizione il modello può essere scritto in questo modo:

$\hat{y}_i = 45291.7 - 71998.4 \cdot \text{tempo} + 31927.1 \cdot \text{tempo}^2 - 2962.2 \cdot \text{tempo}^3$ per FL

$\hat{y}_i = (45291.7 - 17121.8) - (71998.4 + 19352.4) \cdot \text{tempo} + (31927.1 - 4965.4) \cdot \text{tempo}^2 - (2962.2 + 371.5) \cdot \text{tempo}^3$ per SH60

$\hat{y}_i = (45291.7 - 11949.5) - (71998.4 + 12389.1) \cdot \text{tempo} + (31927.1 - 2744.4) \cdot \text{tempo}^2 - (2962.2 + 172.1) \cdot \text{tempo}^3$ per HSH60

$\hat{y}_i = (45291.7 - 10061.6) - (71998.4 + 12252.8) \cdot \text{tempo} + (31927.1 - 3163.4) \cdot \text{tempo}^2 - (2962.2 + 300.5) \cdot \text{tempo}^3$ per SH30

Dai risultati si vede che nessun termine di interazione risulta significativo; dunque, il modello può essere semplificato eliminando questi termini.

I coefficienti che risultano statisticamente significativi sono quelli relativi all'intercetta, alla variabile tempo e alle sue potenze, riferiti alla curva FL.

Il modello, dunque, suggerisce che non vi è una differenza significativa nella stima della curva di regressione tra la condizione FL e le altre.

Questo si può vedere anche nel grafico con le curve di regressione stimate riportato in Figura 3.27.

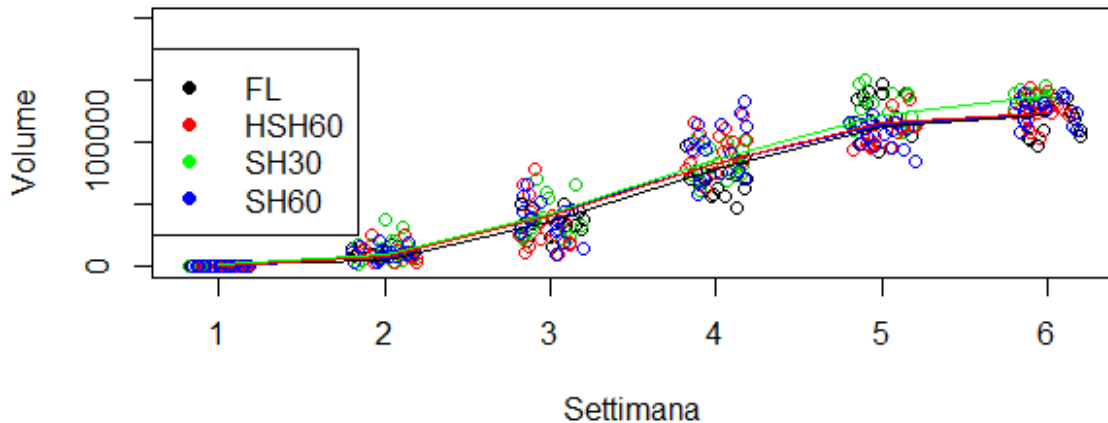


Figura 3.27. Diagramma di dispersione tra tempo e volume per le patate e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Come si può notare dal grafico le curve stimate hanno tutte lo stesso andamento e sono quasi coincidenti.

Successivamente viene presa come variabile dipendente per il modello di regressione il numero di foglie, i cui risultati sono:

```
Call:
lm(formula = nfoglie ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

```
Residuals:
    Min     1Q   Median     3Q     Max
-31.578  -4.459   0.351   3.973  56.604
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -102.7500     8.5793  -11.976 < 2e-16 ***
d2             39.5000    12.1330   3.256 0.001237 **
d3             29.0000    12.1330   2.390 0.017342 *
d4             37.1667    12.1330   3.063 0.002350 **
tempo        132.3514     9.7753  13.539 < 2e-16 ***
tempo2       -27.5074     3.1281  -8.794 < 2e-16 ***
tempo3         1.8733     0.2956   6.337 6.84e-10 ***
d2:tempo     -61.6749    13.8244  -4.461 1.08e-05 ***
d2:tempo2    20.2321     4.4237   4.574 6.56e-06 ***
d2:tempo3    -1.8698     0.4180  -4.473 1.03e-05 ***
d3:tempo     -45.1972    13.8244  -3.269 0.001179 **
d3:tempo2    14.7545     4.4237   3.335 0.000939 ***
d3:tempo3    -1.3698     0.4180  -3.277 0.001150 **
d4:tempo     -57.6129    13.8244  -4.167 3.84e-05 ***
d4:tempo2    19.3127     4.4237   4.366 1.65e-05 ***
d4:tempo3    -1.8154     0.4180  -4.343 1.82e-05 ***
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 9.518 on 368 degrees of freedom
 Multiple R-squared: 0.9433, Adjusted R-squared: 0.941
 F-statistic: 408.4 on 15 and 368 DF, p-value: < 2.2e-16

La stima risulta: $\hat{y}_i = -102.7500 + 39.5000*d2 + 29.0000*d3 + 37.1667*d4 + 132.3514*tempo - 27.5074*tempo^2 + 1.8733*tempo^3 - 61.6749*d2*tempo + 20.2321*d2*tempo^2 - 1.8698*d2*tempo^3 - 45.1972*d3*tempo + 14.7545*d3*tempo^2 - 1.3698*d3*tempo^3 - 57.6129*d4*tempo + 19.3127*d4*tempo^2 - 1.8154*d4*tempo^3$

Tutte le variabili risultano significative, quindi si rifiuta l'ipotesi di nullità di tutti i coefficienti. In pratica esiste una differenza nella stima dell'equazione di regressione di FL rispetto a SH60, HSH60 e SH30.

Il coefficiente di determinazione R^2 indica che il 94.33% della variabilità della risposta viene spiegata dal modello.

Il relativo grafico è mostrato in Figura 3.28.

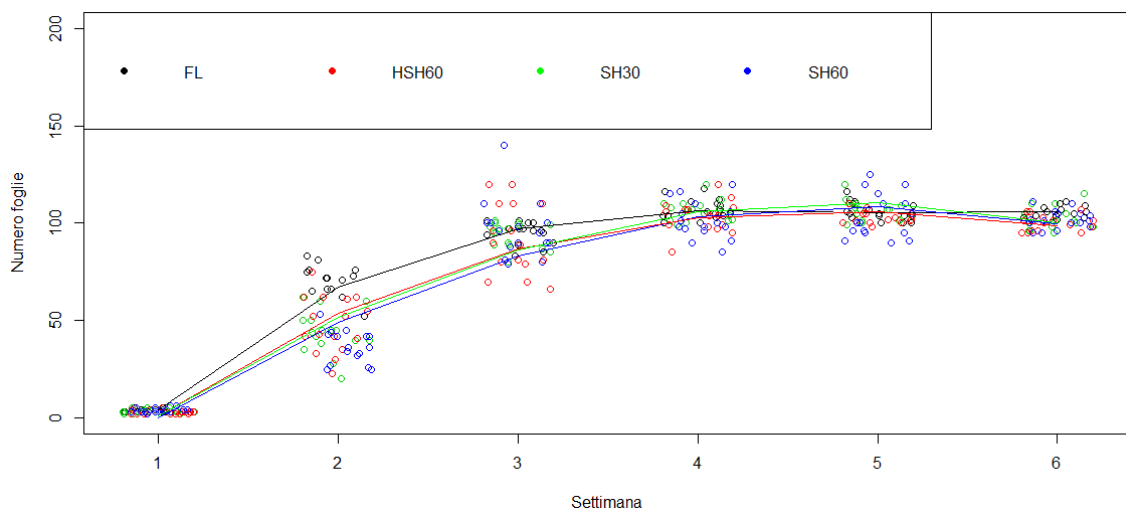


Figura 3.28. Diagramma di dispersione tra tempo e numero di foglie per le patate e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Dopodiché si è presa in esame “lenght_leaves”, ossia la lunghezza delle foglie, come variabile risposta.

Il modello stimato ha fornito i seguenti risultati:

```
Call:
lm(formula = lfoglie ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.7192	-0.7059	-0.0913	0.7460	5.9469

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-3.85417	1.19022	-3.238	0.001312	**
d2	-0.52083	1.68322	-0.309	0.757172	
d3	0.77083	1.68322	0.458	0.647257	
d4	-0.52083	1.68322	-0.309	0.757172	
tempo	7.52687	1.35614	5.550	5.47e-08	***
tempo2	-1.47644	0.43396	-3.402	0.000742	***

```

tempo3      0.09491    0.04101    2.314  0.021197 *
d2:tempo    1.53985    1.91787    0.803  0.422554
d2:tempo2  -0.81969    0.61371   -1.336  0.182494
d2:tempo3    0.10475    0.05799    1.806  0.071712 .
d3:tempo   -0.30696    1.91787   -0.160  0.872927
d3:tempo2  -0.05977    0.61371   -0.097  0.922466
d3:tempo3    0.01505    0.05799    0.259  0.795436
d4:tempo    1.22561    1.91787    0.639  0.523189
d4:tempo2  -0.52207    0.61371   -0.851  0.395497
d4:tempo3    0.06481    0.05799    1.118  0.264460
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.32 on 368 degrees of freedom
Multiple R-squared:  0.7745, Adjusted R-squared:  0.7653
F-statistic: 84.25 on 15 and 368 DF,  p-value: < 2.2e-16

```

La stima del modello è: $\hat{y}_i = -3.85417 - 0.52083 \cdot d2 + 0.77083 \cdot d3 - 0.52083 \cdot d4 + 7.52687 \cdot \text{tempo} - 1.47644 \cdot \text{tempo}^2 + 0.09491 \cdot \text{tempo}^3 + 1.53985 \cdot d2 \cdot \text{tempo} - 0.81969 \cdot d2 \cdot \text{tempo}^2 + 0.10475 \cdot d2 \cdot \text{tempo}^3 - 0.30696 \cdot d3 \cdot \text{tempo} - 0.05977 \cdot d3 \cdot \text{tempo}^2 + 0.01505 \cdot d3 \cdot \text{tempo}^3 + 1.22561 \cdot d4 \cdot \text{tempo} - 0.52207 \cdot d4 \cdot \text{tempo}^2 + 0.06481 \cdot d4 \cdot \text{tempo}^3$

I coefficienti dei termini di interazione risultano tutti non significativi; dunque, si accetta l'ipotesi di nullità di questi coefficienti e le relative variabili vengono eliminate dal modello.

Risultano statisticamente significative le variabili relative al tempo e l'intercetta.

Il modello stimato spiega il 77.45% della variabilità della risposta.

Il grafico delle curve stimate dal modello viene riportato di seguito in Figura 3.29.

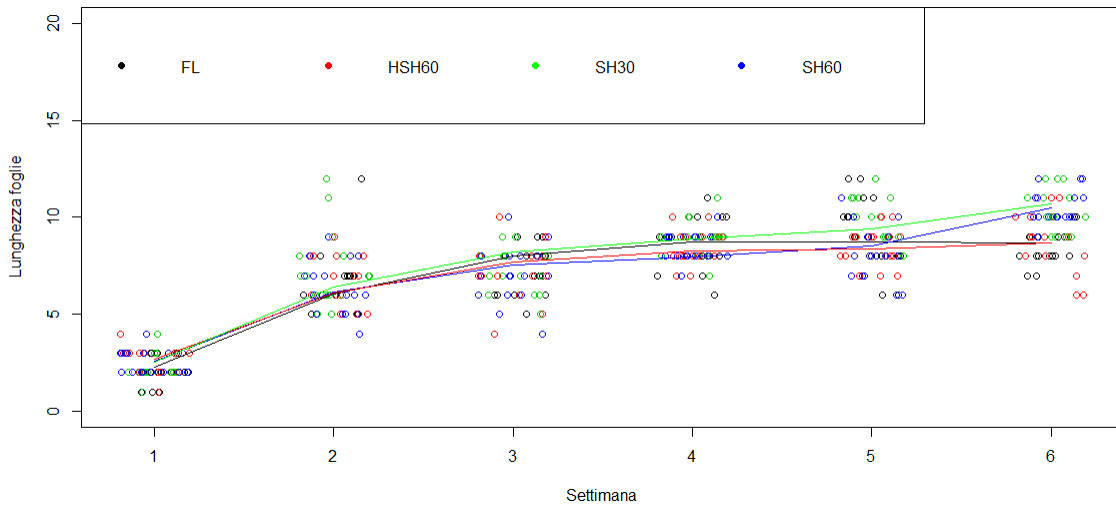


Figura 3.29. Diagramma di dispersione tra tempo e lunghezza delle foglie per le patate e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

Infine, si è presa come variabile risposta nel modello di regressione il numero di fiori, i cui risultati sono:

```

Call:
lm(formula = nfiori ~ d2 + d3 + d4 + tempo + tempo2 + tempo3 +
    d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
    d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3)

```

Residuals:
 Min 1Q Median 3Q Max
 -4.7927 -0.7758 -0.0233 0.3710 4.8194

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	3.958333	1.293421	3.060	0.00237	**
d2	-0.354167	1.829174	-0.194	0.84658	
d3	-0.812500	1.829174	-0.444	0.65717	
d4	1.895833	1.829174	1.036	0.30068	
tempo	-6.217923	1.473732	-4.219	3.09e-05	***
tempo2	2.565972	0.471587	5.441	9.68e-08	***
tempo3	-0.269676	0.044564	-6.051	3.54e-09	***
d2:tempo	0.095569	2.084172	0.046	0.96345	
d2:tempo2	0.201389	0.666925	0.302	0.76285	
d2:tempo3	-0.037037	0.063023	-0.588	0.55711	
d3:tempo	0.822173	2.084172	0.394	0.69345	
d3:tempo2	-0.084821	0.666925	-0.127	0.89887	
d3:tempo3	-0.005208	0.063023	-0.083	0.93418	
d4:tempo	-3.043816	2.084172	-1.460	0.14502	
d4:tempo2	1.403026	0.666925	2.104	0.03608	*
d4:tempo3	-0.160301	0.063023	-2.544	0.01138	*

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.435 on 368 degrees of freedom
 Multiple R-squared: 0.6076, Adjusted R-squared: 0.5916
 F-statistic: 37.99 on 15 and 368 DF, p-value: < 2.2e-16

Solo i termini di interazione $d4 \times tempo^2$ e $d4 \times tempo^3$ risultano significativi, indicando che la curva per la condizione SH30 è differente rispetto le altre. Invece i restanti coefficienti possono essere eliminati dal modello.

L'adattamento del modello è discreto, in quanto il coefficiente R^2 è pari a 0.6076

Il grafico delle curve stimate viene riportato in Figura 3.30.

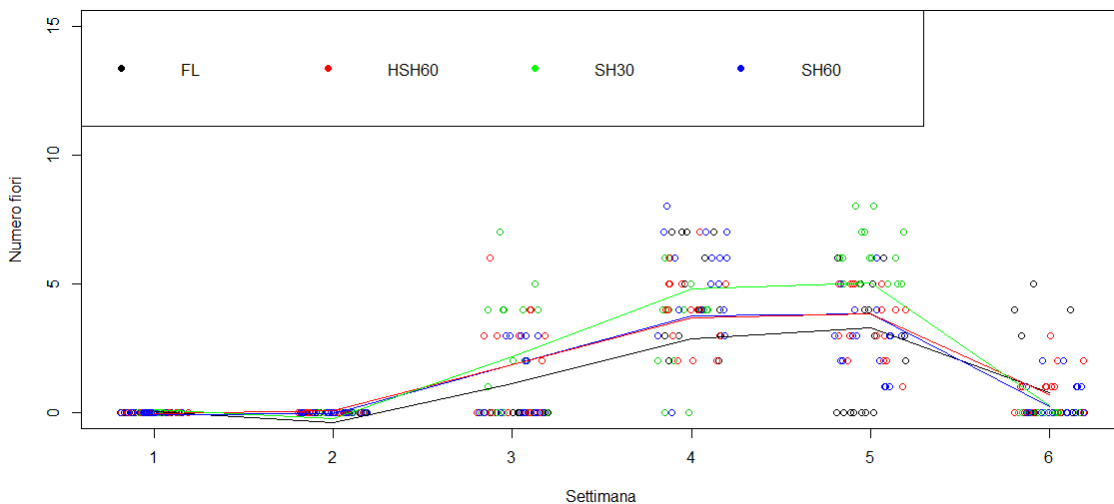


Figura 3.30. Diagramma di dispersione tra tempo e numero di fiori per le patate e curve di regressione stimate con trend cubico per ciascuna condizione (FL-HSH60-SH30-SH60)

3.3.3 Regressione quantilica

Nei prossimi modelli verranno considerate come variabili risposte “volume”, “number_leaves”, “length_leaves” e “number_flowers”.

Come variabili indipendenti “week” e “light”. Per inserire la variabile “light” nel modello sono state create quattro variabili dummy così definite:

$$x_{i1} = d1 = \begin{cases} 1 & \text{se light} = \text{FL} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i2} = d2 = \begin{cases} 1 & \text{se light} = \text{SH60} \\ 0 & \text{altrimenti} \end{cases}$$

$$x_{i3} = d3 = \begin{cases} 1 & \text{se light} = \text{HSH60} \\ 0 & \text{altrimenti} \end{cases} \quad x_{i4} = d4 = \begin{cases} 1 & \text{se light} = \text{SH30} \\ 0 & \text{altrimenti} \end{cases}$$

Come per i modelli precedenti, si è considerata come variabile di riferimento d1 e i modelli sono stati stimati per il quantile di ordine 0.10 e 0.90.

I risultati relativi ai modelli con variabile risposta volume sono:

```
Call: rq(formula = volume ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.1
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	36392.60000	12977.60517	2.80426	0.00531
tempo	-58809.36667	18058.24229	-3.25665	0.00123
tempo2	24621.90000	6375.52563	3.86194	0.00013
tempo3	-2155.13333	626.28864	-3.44112	0.00065
d2	-2854.60000	16003.93517	-0.17837	0.85853
d3	24447.65000	18612.04930	1.31354	0.18982
d4	21333.90000	17361.01635	1.22884	0.21992
tempo:d2	2958.36667	22033.06303	0.13427	0.89326
tempo2:d2	-79.40000	7813.29545	-0.01016	0.99190
tempo3:d2	-27.36667	776.54602	-0.03524	0.97191
tempo:d3	-35083.00833	26167.43244	-1.34071	0.18084
tempo2:d3	11691.35000	9180.04594	1.27356	0.20362
tempo3:d3	-1071.99167	910.01929	-1.17799	0.23956
tempo:d4	-31562.88333	24119.75623	-1.30859	0.19149
tempo2:d4	11224.10000	8359.24045	1.34272	0.18019
tempo3:d4	-986.11667	812.30758	-1.21397	0.22554

```
Call: rq(formula = volume ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.9
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	40595.50000	8902.77425	4.55987	0.00001
tempo	-74178.85000	11809.51186	-6.28128	0.00000
tempo2	37464.65000	4480.72138	8.36130	0.00000
tempo3	-3740.30000	477.48027	-7.83341	0.00000
d2	-12353.30000	19372.74025	-0.63766	0.52409
d3	-38154.70000	20549.81370	-1.85669	0.06415
d4	-36449.50000	15156.19005	-2.40492	0.01667
tempo:d2	14840.58333	27326.60657	0.54308	0.58740
tempo2:d2	-2525.85000	10302.86841	-0.24516	0.80647
tempo3:d2	67.56667	1059.09708	0.06380	0.94917
tempo:d3	53348.25000	28276.60916	1.88666	0.05999
tempo2:d3	-16705.78333	9683.42961	-1.72519	0.08533
tempo3:d3	1488.23333	940.04800	1.58315	0.11425
tempo:d4	50801.31667	20203.39909	2.51449	0.01235

tempo2:d4	-15801.55000	7012.74188	-2.25326	0.02483
tempo3:d4	1421.73333	699.48812	2.03253	0.04282

Per il quantile 0.10 risultano significative l'intercetta e le variabili temporali, mentre non lo risultano le dummy e i termini di interazione. Può dunque essere accettata l'ipotesi di nullità di questi coefficienti, indicando che non vi sono differenze significative nella stima tra le diverse condizioni per le piante con un valore basso del volume.

Per il quantile 0.90 risultano significative l'intercetta, le variabili temporali, la variabile dicotomica d4 associata alla condizione SH30 e i termini di interazione relativi ad essa.

Le curve di regressione quantilica stimate sono riportate in Figura 3.31.

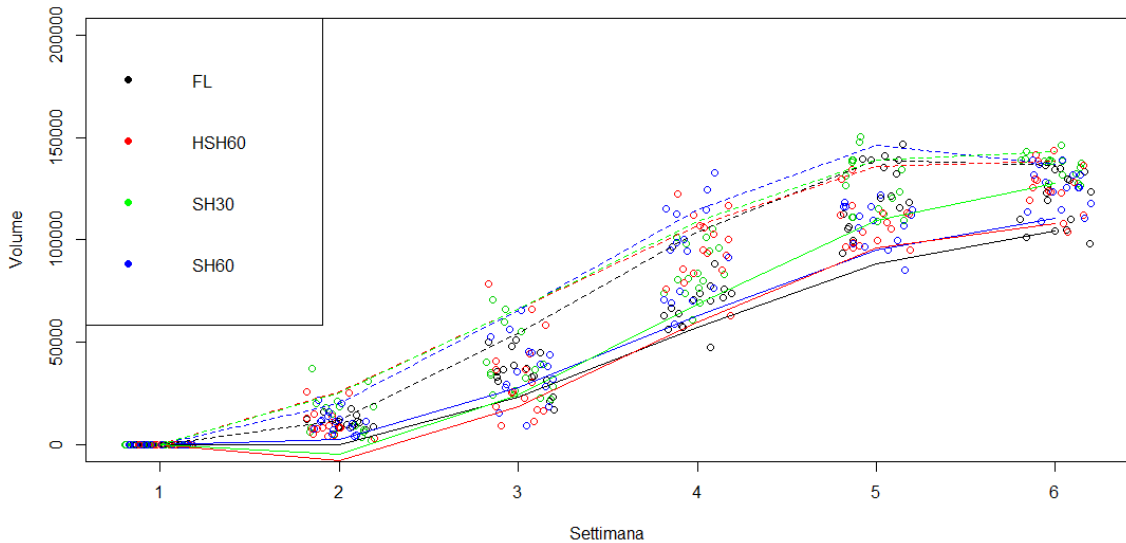


Figura 3.31. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta volume

Successivamente si è considerata come variabile dipendente “number_leaves”, ossia il numero di foglie della pianta, i cui modelli hanno portato ai seguenti risultati:

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

tau: [1] 0.1

Coefficients:

	Value	std. Error	t value	Pr(> t)
(Intercept)	-85.00000	8.56582	-9.92316	0.00000
tempo	106.16667	11.56488	9.18009	0.00000
tempo2	-20.50000	3.70425	-5.53418	0.00000
tempo3	1.33333	0.34000	3.92159	0.00010
d2	58.50000	12.41493	4.71207	0.00000
d3	56.00000	12.63521	4.43206	0.00001
d4	49.50000	12.74693	3.88329	0.00012
tempo:d2	-79.61667	16.23988	-4.90254	0.00000
tempo2:d2	23.05000	5.09660	4.52262	0.00001
tempo3:d2	-1.93333	0.46880	-4.12399	0.00005
tempo:d3	-77.70000	16.84236	-4.61337	0.00001
tempo2:d3	23.80000	5.26156	4.52338	0.00001
tempo3:d3	-2.10000	0.47527	-4.41856	0.00001
tempo:d4	-69.11667	16.69542	-4.13986	0.00004
tempo2:d4	21.55000	5.16265	4.17421	0.00004
tempo3:d4	-1.93333	0.46660	-4.14342	0.00004

```
Call: rq(formula = nfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.9
```

```
Coefficients:
```

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-119.00000	6.79335	-17.51714	0.00000
tempo	155.16667	8.82752	17.57760	0.00000
tempo2	-33.50000	2.78132	-12.04464	0.00000
tempo3	2.33333	0.25503	9.14908	0.00000
d2	40.25000	16.75282	2.40258	0.01677
d3	-10.00000	15.40858	-0.64899	0.51675
d4	27.00000	10.41060	2.59351	0.00988
tempo:d2	-57.70833	23.87880	-2.41672	0.01615
tempo2:d2	19.25000	8.03840	2.39476	0.01713
tempo3:d2	-1.79167	0.76175	-2.35203	0.01920
tempo:d3	12.40000	21.25835	0.58330	0.56005
tempo2:d3	-3.70000	6.66656	-0.55501	0.57923
tempo3:d3	0.30000	0.59877	0.50103	0.61665
tempo:d4	-37.50000	14.22124	-2.63690	0.00872
tempo2:d4	11.50000	4.70843	2.44243	0.01506
tempo3:d4	-1.00000	0.44754	-2.23444	0.02605

Per il quantile di ordine 0.10 risultano significativi i coefficienti di tutte le variabili.

Per il quantile di ordine 0.90 invece non risultano significativi i coefficienti di d3 e dei termini di interazione associati. Il grafico relativo è riportato in Figura 3.32.

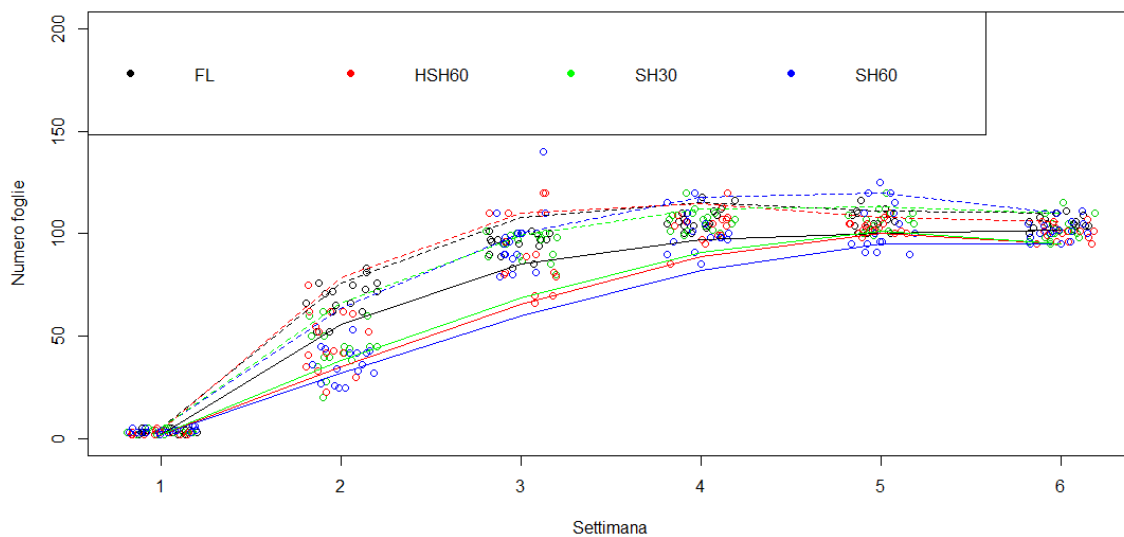


Figura 3.32. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta numero di foglie

Dopodiché sono stati stimati i modelli con variabile risposta “length_leaves”, che hanno prodotto i seguenti risultati:

```
Call: rq(formula = lfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.1
```

```
Coefficients:
```

Value	Std. Error	t value	Pr(> t)
-------	------------	---------	----------

(Intercept)	-6.25000	2.02100	-3.09253	0.00214
tempo	9.40833	2.81826	3.33835	0.00093
tempo2	-2.35000	1.01071	-2.32510	0.02061
tempo3	0.19167	0.10019	1.91300	0.05652
d2	1.75000	2.30673	0.75865	0.44855
d3	5.50000	2.51264	2.18894	0.02923
d4	7.00000	2.57669	2.71666	0.00691
tempo:d2	-0.69167	3.17147	-0.21809	0.82748
tempo2:d2	-0.10000	1.12920	-0.08856	0.92948
tempo3:d2	0.04167	0.11186	0.37248	0.70975
tempo:d3	-6.48333	3.48122	-1.86237	0.06335
tempo2:d3	2.20000	1.23102	1.78713	0.07474
tempo3:d3	-0.21667	0.12209	-1.77471	0.07677
tempo:d4	-8.63333	3.60415	-2.39539	0.01710
tempo2:d4	2.90000	1.24974	2.32049	0.02086
tempo3:d4	-0.26667	0.12140	-2.19660	0.02867

```
Call: rq(formula = lfoglie ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

tau: [1] 0.9

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	-5.00000	1.70447	-2.93346	0.00356
tempo	9.70000	2.09335	4.63371	0.00000
tempo2	-1.80000	0.65299	-2.75657	0.00613
tempo3	0.10000	0.06182	1.61760	0.10660
d2	-2.00000	2.27055	-0.88084	0.37898
d3	-1.60000	2.15869	-0.74119	0.45905
d4	2.00000	3.48798	0.57340	0.56673
tempo:d2	3.36667	2.90118	1.16045	0.24662
tempo2:d2	-1.55000	0.92702	-1.67203	0.09537
tempo3:d2	0.18333	0.08755	2.09398	0.03695
tempo:d3	2.68333	2.73067	0.98266	0.32642
tempo2:d3	-1.22500	0.87490	-1.40016	0.16231
tempo3:d3	0.14167	0.08392	1.68815	0.09223
tempo:d4	-0.90000	3.65190	-0.24645	0.80547
tempo2:d4	-0.15000	1.06156	-0.14130	0.88771
tempo3:d4	0.05000	0.09464	0.52834	0.59758

Risultano significativi per il quantile di ordine 0.10 l'intercetta, il tempo, il tempo², le dummy d3 e d4 e i termini di interazione tra la variabile temporale e quella dicotomica d4.

Invece per il quantile 0.90 risultano significativi l'intercetta, il tempo e il tempo².

Il grafico con le curve stimate è riportato in Figura 3.33.

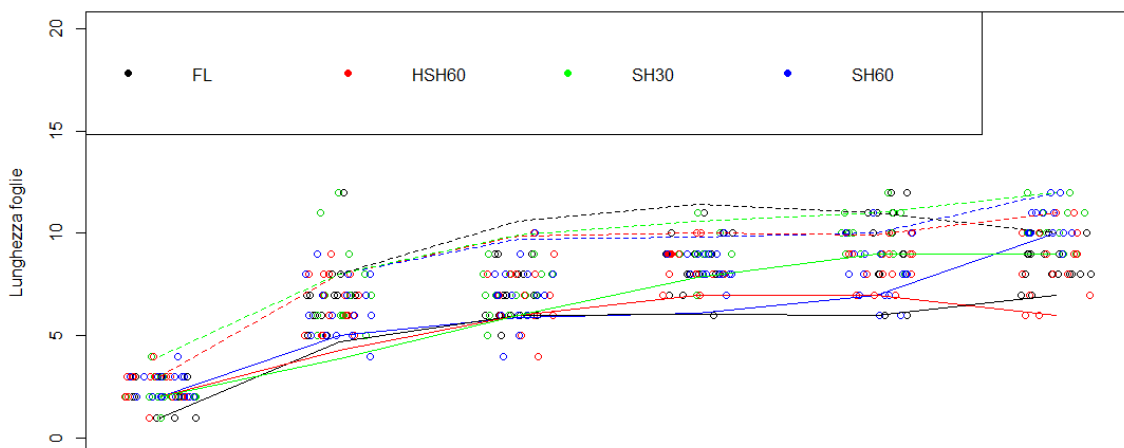


Figura 3.33. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta lunghezza delle foglie

Infine, sono stati i modelli per la variabile risposta “number_flowers”, che hanno generato i seguenti risultati:

```
Call: rq(formula = nfiori ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.1
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	0.00000	0.93841	0.00000	1.00000
tempo	0.00000	1.25429	0.00000	1.00000
tempo2	0.00000	0.46955	0.00000	1.00000
tempo3	0.00000	0.04837	0.00000	1.00000
d2	2.25000	1.39766	1.60984	0.10829
d3	4.50000	1.39506	3.22568	0.00137
d4	6.00000	2.46898	2.43015	0.01557
tempo:d2	-3.37500	1.89041	-1.78533	0.07503
tempo2:d2	1.25000	0.71228	1.75493	0.08010
tempo3:d2	-0.12500	0.07362	-1.69794	0.09036
tempo:d3	-6.75000	1.86116	-3.62677	0.00033
tempo2:d3	2.50000	0.67501	3.70364	0.00025
tempo3:d3	-0.25000	0.06833	-3.65885	0.00029
tempo:d4	-9.00000	3.54572	-2.53827	0.01155
tempo2:d4	3.33333	1.31976	2.52571	0.01197
tempo3:d4	-0.33333	0.13335	-2.49960	0.01287

```
Call: rq(formula = nfiori ~ tempo + tempo2 + tempo3 + d2 + d3 + d4 +
  d2:tempo + d2:tempo2 + d2:tempo3 + d3:tempo + d3:tempo2 +
  d3:tempo3 + d4:tempo + d4:tempo2 + d4:tempo3, tau = taus)
```

```
tau: [1] 0.9
```

Coefficients:

	Value	Std. Error	t value	Pr(> t)
(Intercept)	5.20000	1.29045	4.02961	0.00007
tempo	-8.60000	1.78459	-4.81903	0.00000
tempo2	3.80000	0.69200	5.49135	0.00000
tempo3	-0.40000	0.07384	-5.41680	0.00000
d2	0.60000	1.75315	0.34224	0.73236
d3	-4.20000	1.71684	-2.44635	0.01490
d4	0.55000	1.58667	0.34664	0.72906
tempo:d2	-1.03333	2.38225	-0.43376	0.66471
tempo2:d2	0.50000	0.89793	0.55683	0.57798
tempo3:d2	-0.06667	0.09364	-0.71193	0.47696
tempo:d3	6.06667	2.32285	2.61173	0.00938
tempo2:d3	-2.05000	0.85073	-2.40971	0.01646
tempo3:d3	0.18333	0.08782	2.08751	0.03753
tempo:d4	-1.35833	2.14070	-0.63453	0.52613
tempo2:d4	0.95000	0.81205	1.16988	0.24281
tempo3:d4	-0.14167	0.08560	-1.65503	0.09877

Per il quantile di ordine 0.10 risultano significative solo le variabili dicotomiche d3 e d4 e i loro termini di interazione con le variabili temporali.

Per il quantile di ordine 0.90 risultano significative l'intercetta, le variabili temporali, la dummy d3 e i relativi termini di interazione.

Il grafico con le curve di regressione quantilica stimate per entrambi i quantili viene riportato qui sotto in Figura 3.34.

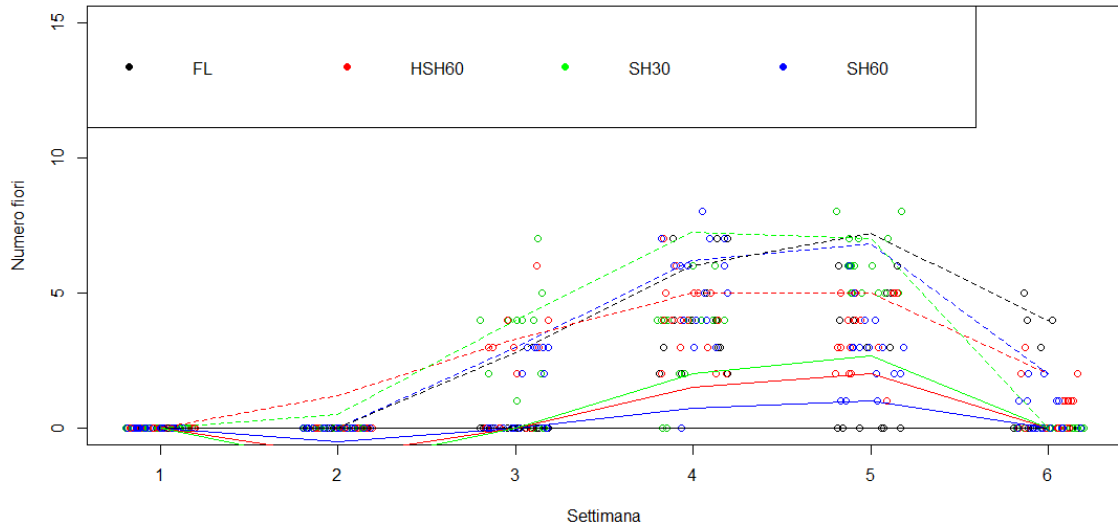


Figura 3.34. Curve di regressione quantilica per il quantile 0.10 (linee continue) e il quantile 0.90 (linee tratteggiate) per ciascuna condizione (FL, SH60, HSH60, SH30) per variabile risposta numero di fiori

CONCLUSIONI

I risultati relativi ai modelli di regressione polinomiale per l'erba medica mostrano che la condizione di piena luce FL risulta leggermente migliore per la crescita media dei parametri delle piante rispetto alla condizione d'ombra SH. I modelli di regressione quantilica, invece, mostrano che per il quantile di ordine 0.10, ossia per i valori bassi della distribuzione della variabile risposta, non vi è una differenza significativa tra le due condizioni FL e SH. Tuttavia, per il quantile 0.90, ossia per valori alti della distribuzione della variabile risposta, risulta nuovamente che la condizione di piena luce porta a valori leggermente superiori dei parametri considerati.

Per quanto riguarda le piante dei fagioli, i modelli di regressione lineare con trend polinomiale hanno evidenziato il fatto che la condizione d'ombra che porta a risultati di crescita media dei parametri più simili alla condizione di piena luce è quella di tipo SH30, ossia un ombreggiamento al 30%. Le condizioni d'ombra di tipo SH60 (ombra al 60%) e HSH60 (metà luce e metà ombra al 60%), invece, portano a valori di crescita dei parametri inferiori, di poco, rispetto alle due precedenti. I modelli di regressione polinomiale mostrano che in generale, per entrambi i quantili, l'andamento temporale dei parametri rilevati risulta piuttosto simile per ciascuna condizione.

Infine, dai risultati relativi ai modelli di regressione con trend polinomiale delle piante di patate, consegue che, per la maggior parte delle variabili analizzate, non vi siano differenze significative nell'andamento temporale tra le diverse condizioni in cui sono state coltivate. I modelli di regressione quantilica mostrano, per il quantile di ordine 0.10, ossia per i valori inferiori della distribuzione della variabile risposta, la condizione d'ombra che porta a valori maggiori dei parametri è la condizione SH30. Per il quantile di ordine 0.90 risulta che le varie condizioni portano a risultati che non differiscono molto tra loro.

In conclusione, la condizione di piena luce FL, porta a dei valori maggiori dei parametri per la coltivazione di erba medica, mentre per le piante di fagioli e patate, le differenze nella crescita per le diverse variabili considerate non sono troppo marcate.

Questo, dunque, unito agli innumerevoli vantaggi indotti dall'uso di un sistema agrivoltaico descritti precedentemente, mi porta a concludere che un impianto di questo tipo sia preferibile a uno tradizionale.

BIBLIOGRAFIA

- [1] Agostini et al, (2021), “*Innovative agrivoltaic systems to produce sustainable energy: An economic and environmental assessment*”
- [2] Amaducci et al, (2018), “*Agrivoltaic systems to optimise land use for electric energy production*”
- [3] Davino et al, (2014), “*Quantile Regression. Theory and Applications*”
- [4] Di Fonzo e Lisi, (2005), “*Serie storiche economiche. Analisi statistiche e applicazioni*”
- [5] Dupraz et al, (2011), “*Combining solar photovoltaic panels and food crops for optimising land use: Towards new agrivoltaic schemes*”
- [6] Dinesh e Pearce, (2015), “*The potential of agrivoltaic systems*”
- [7] Giorgio Ghiringhelli, (2018), “*L’agenda 2030 e gli obiettivi di sviluppo sostenibile*”
- [8] Grigoletto et al, (2017), “*Modello lineare. Teoria e applicazioni con R*”
- [9] Marrou et al, (2013), “*Microclimate under agrivoltaic systems: Is crop growth rate affected in the partial shade of solar panels?*”
- [10] Mirandola, (2019), “*Slide del Corso di macchine ed impianti per la produzione di energia*”
- [11] Toledo e Scognamiglio, (2021), “*Agrivoltaic Systems Design and Assessment: A Critical Review, and a Descriptive Model towards a Sustainable Landscape Vision (Three-Dimensional Agrivoltaic Patterns)*”
- [12] Touil, et al, (2021), “*Shading effect of photovoltaic panels on horticulture crops production: a mini review*”
- [13] Weselek et al, (2019), “*Agrophotovoltaic systems: applications, challenges, and opportunities. A review*”
- [14] <http://www.wikipedia.com>

APPENDICE

Grafici erba medica

Vengono qui riportati i grafici dell'andamento del numero di foglie nel corso delle settimane, per ciascuna area considerata.

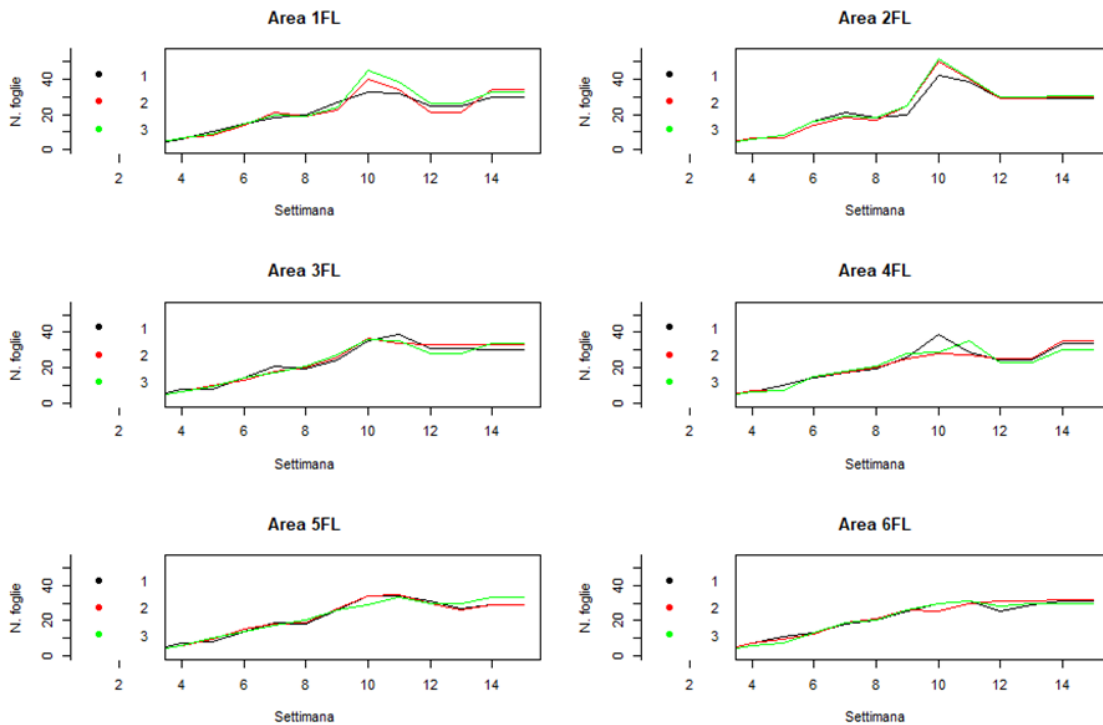


Figura 1. Grafici dell'andamento del numero di foglie per erba medica per le aree FL

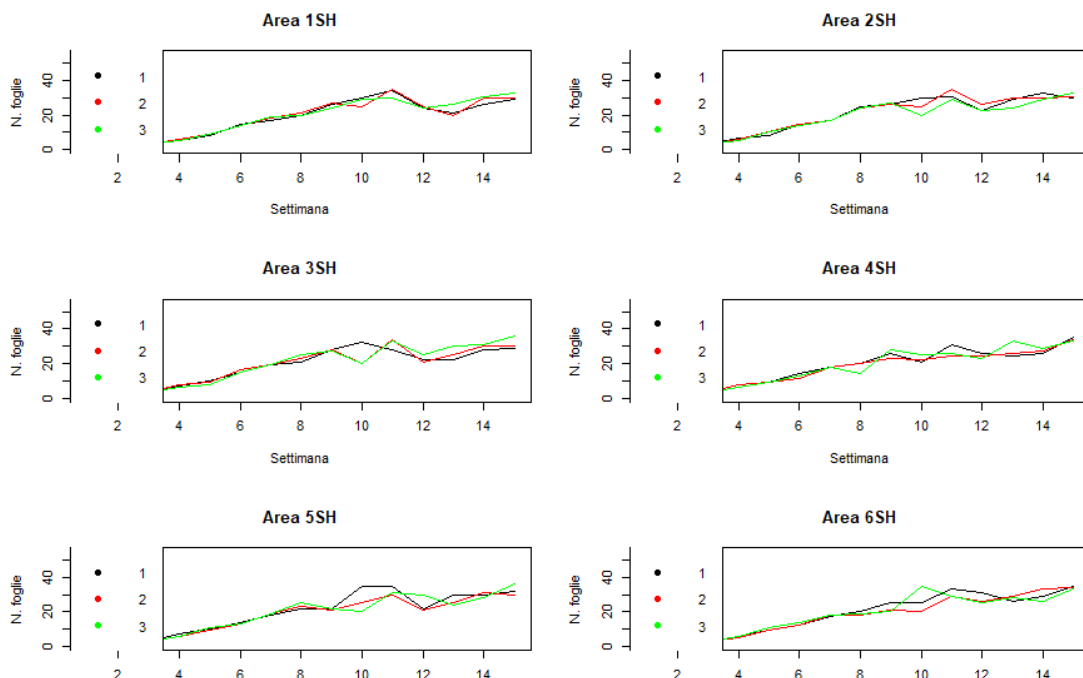


Figura 2. Grafici dell'andamento del numero di foglie per erba medica per le aree 1SH-6SH

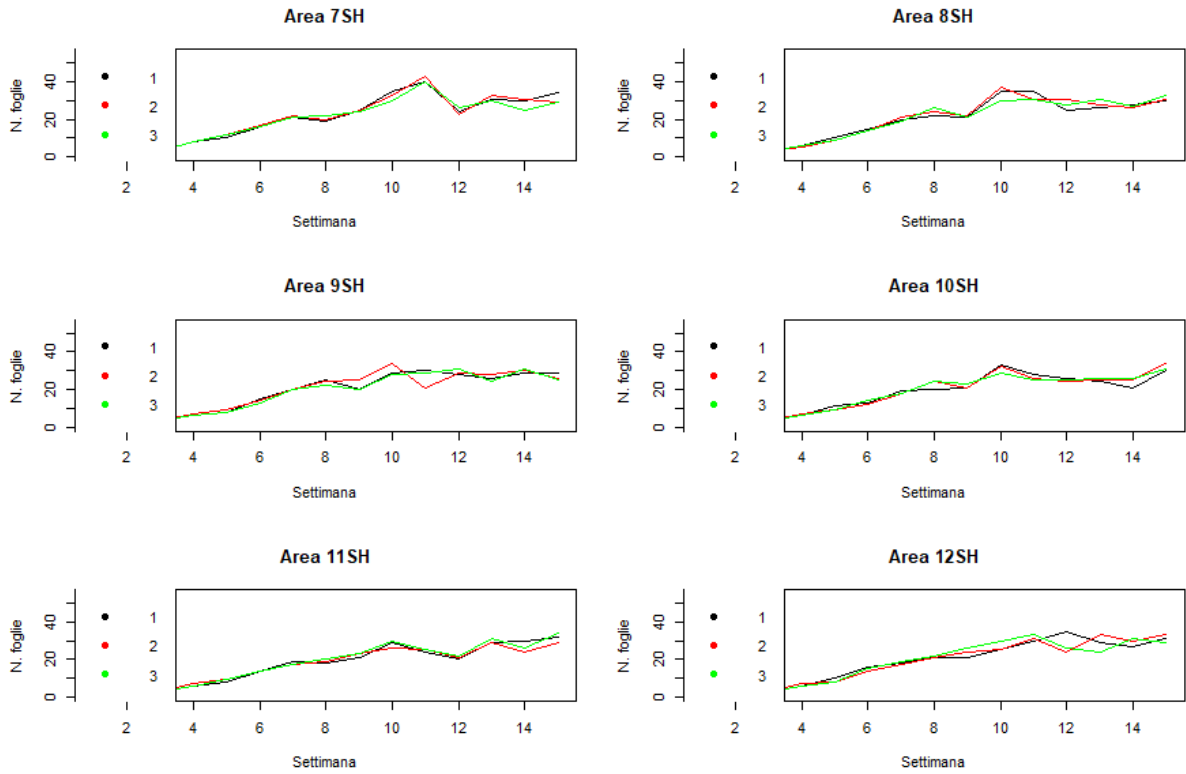


Figura 3. Grafici dell'andamento del numero di foglie per erba medica per le aree 7SH-12SH

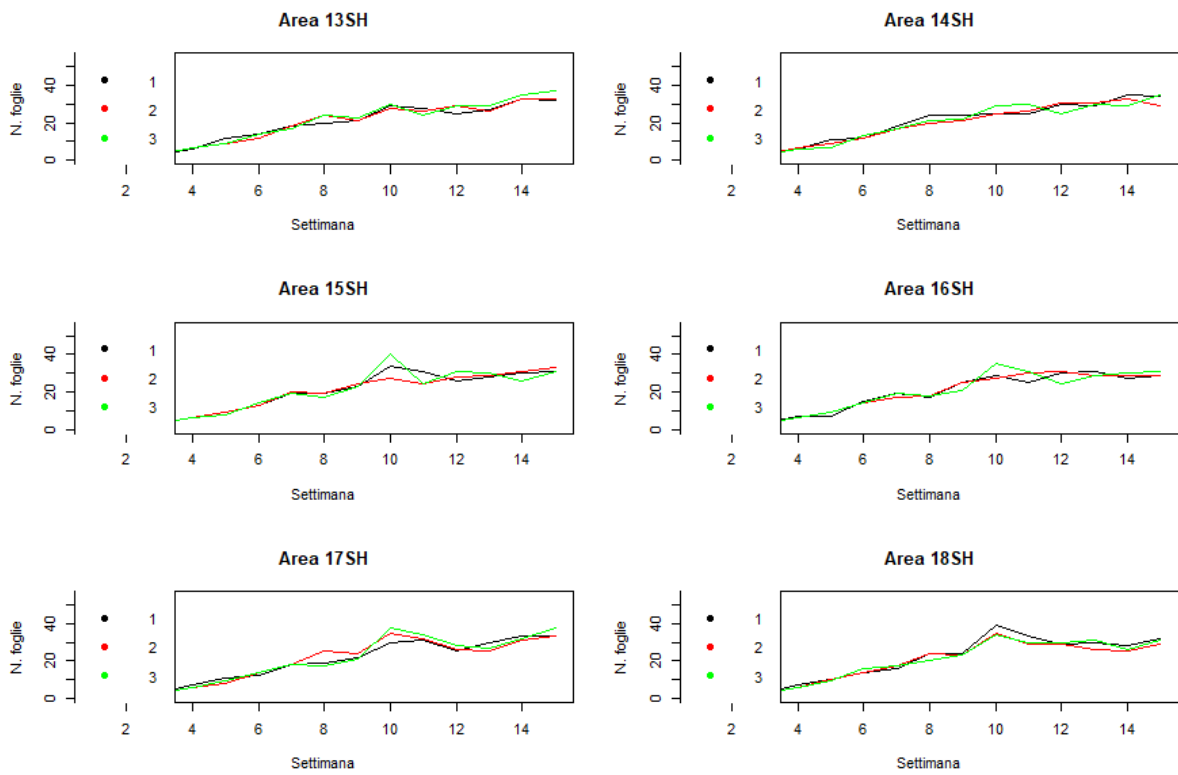


Figura 4. Grafici dell'andamento del numero di foglie per erba medica per le aree 13SH-18SH

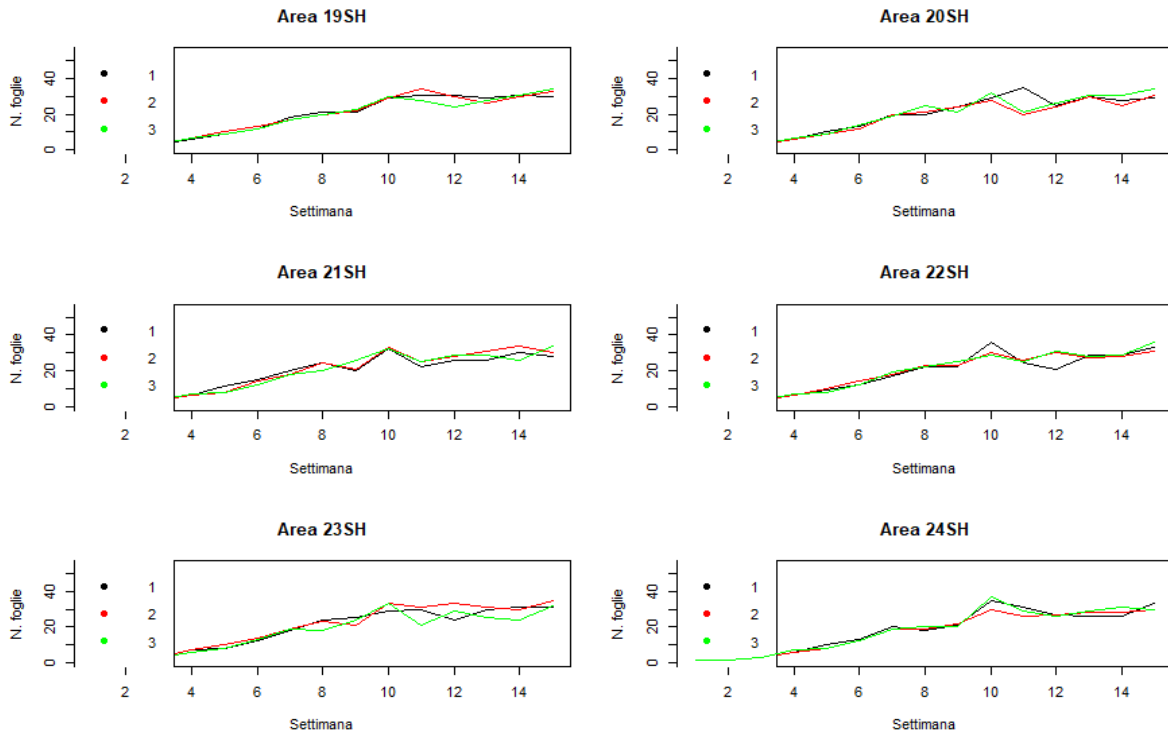


Figura 5. Grafici dell'andamento del numero di foglie per erba medica per le aree 19SH-24SH

Grafici fagioli

Si riportano qui i grafici dell'andamento delle variabili “length_leaves”, “number_flowers” e “number_fruits”, nel corso delle settimane, per le diverse aree considerate (SH60, HSH60, SH30 e FL)

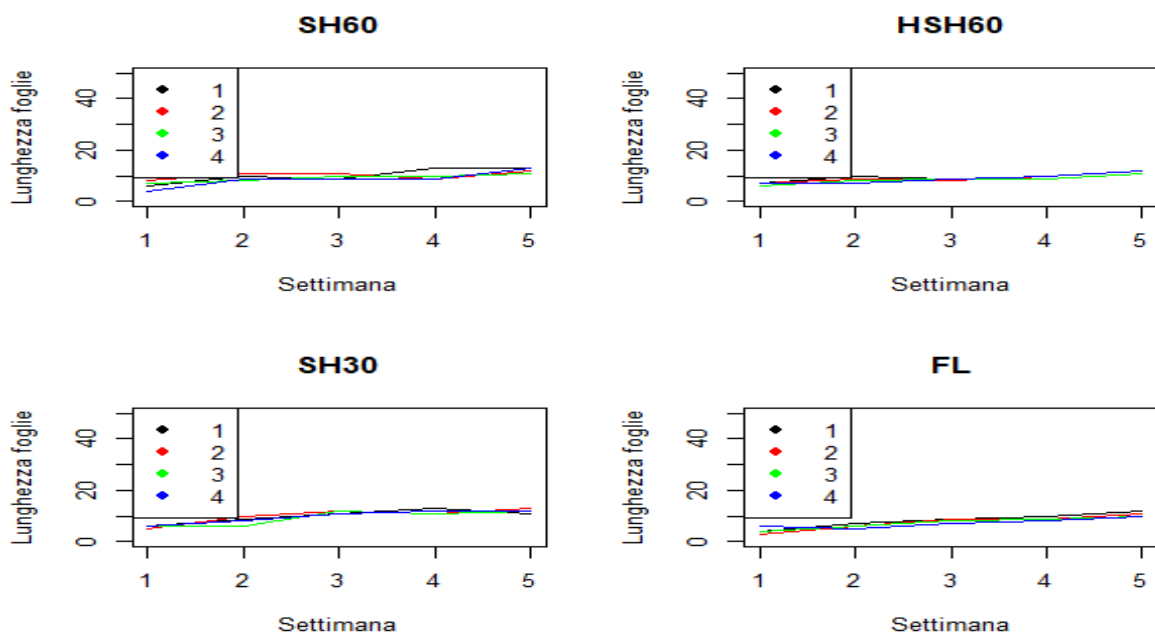


Figura 6. Grafici dell'andamento della lunghezza delle foglie per i fagioli (SH60-HSH60-SH30-FL)

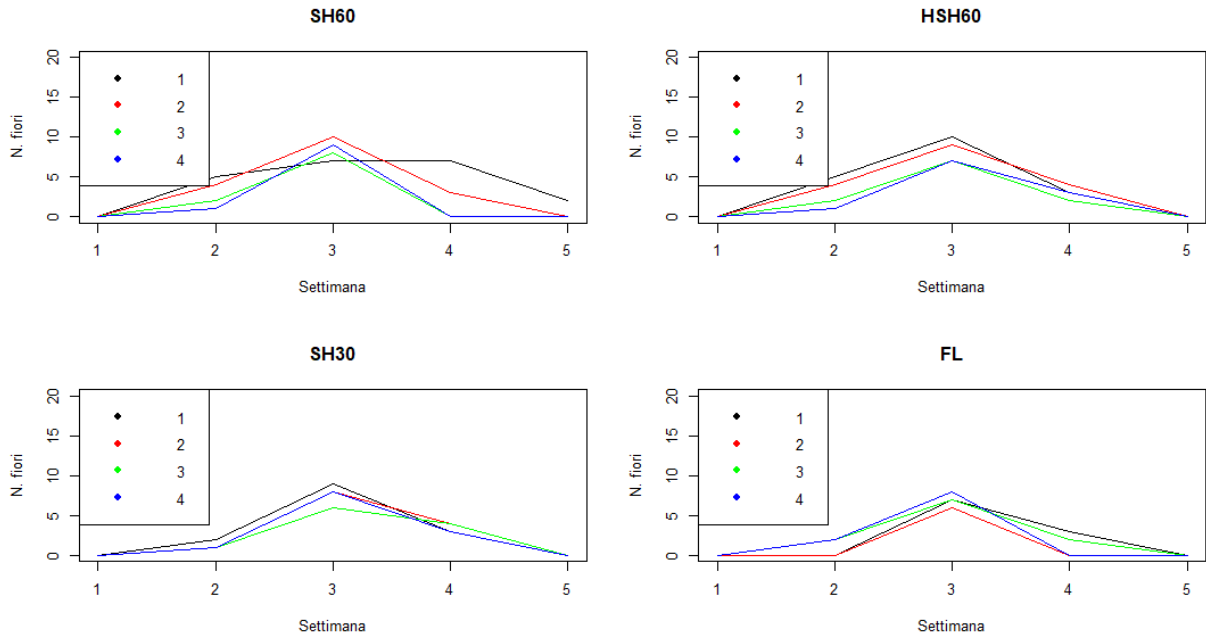


Figura 7. Grafici dell'andamento del numero di fiori per i fagioli (SH60-HSH60-SH30-FL)

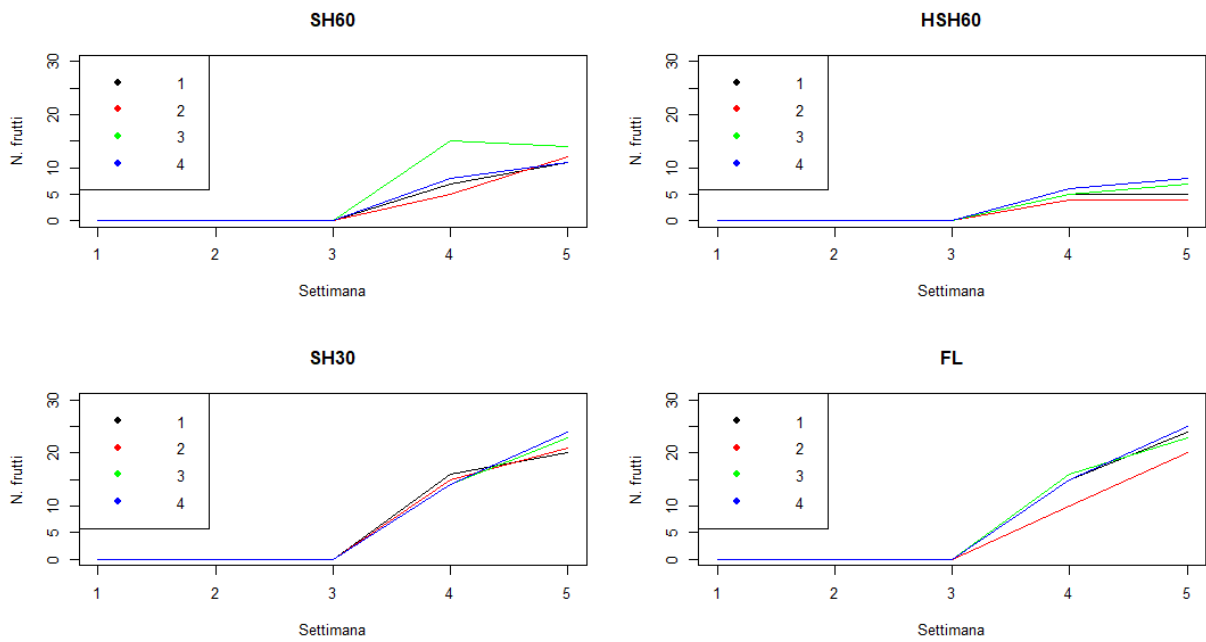


Figura 8. Grafici dell'andamento del numero di frutti per i fagioli (SH60-HSH60-SH30-FL)

Grafici patate

Di seguito sono riportati i grafici dell'andamento del numero di foglie, lunghezza delle foglie e numero di fiori, nel corso delle settimane per le diverse condizioni considerate (FL, SH30, HSH60, SH60)

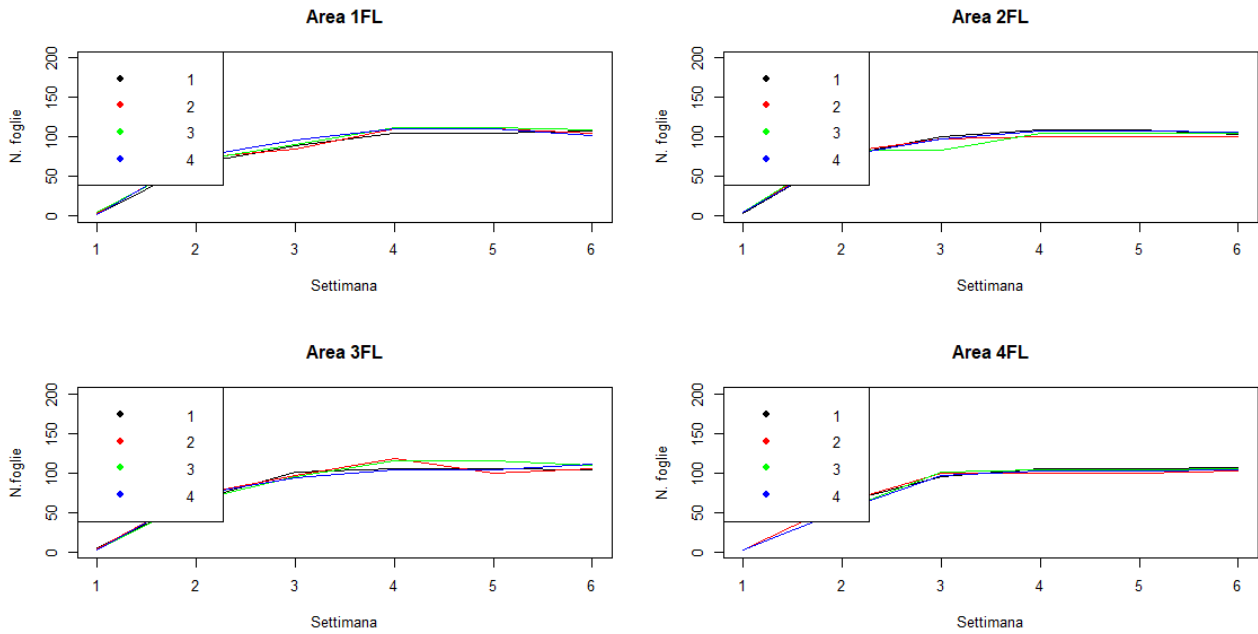


Figura 9. Grafici dell'andamento del numero delle foglie per le patate per aree FL

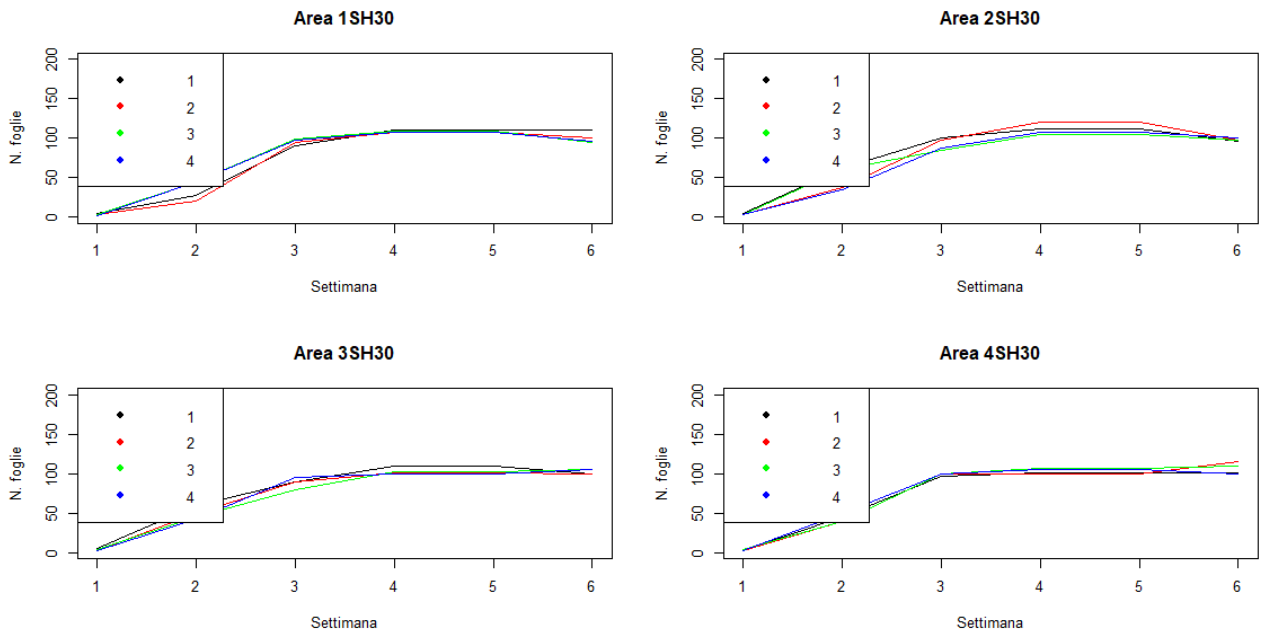


Figura 10. Grafici dell'andamento del numero delle foglie per le patate per aree SH30

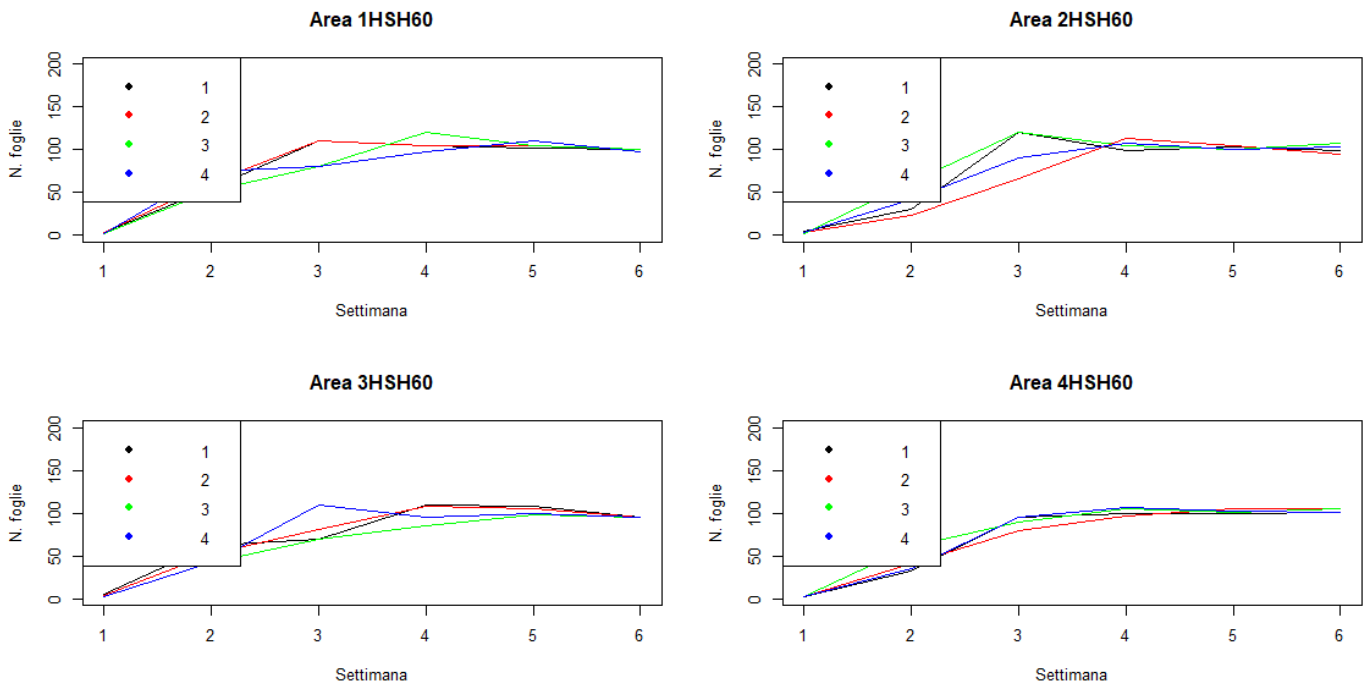


Figura 11. Grafici dell'andamento del numero delle foglie per le patate per aree HSH60

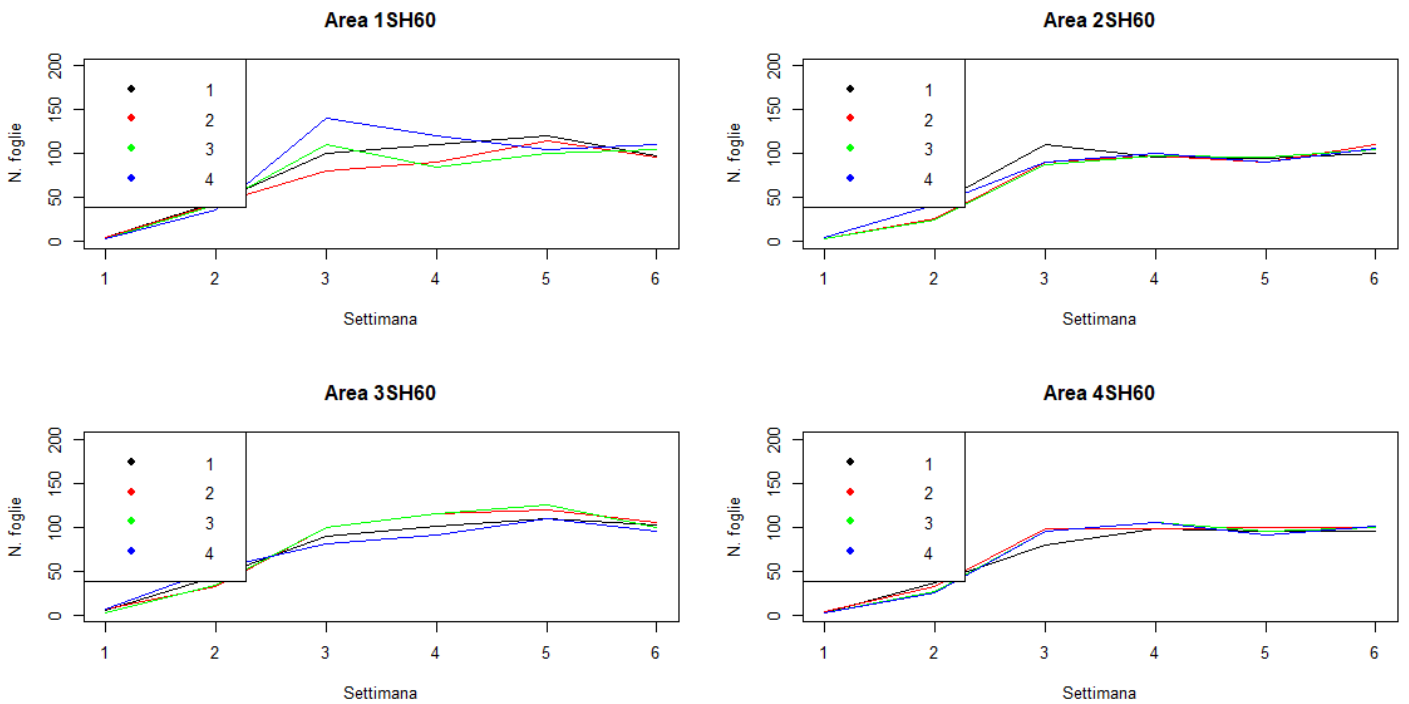


Figura 12. Grafici dell'andamento del numero delle foglie per le patate per aree SH60

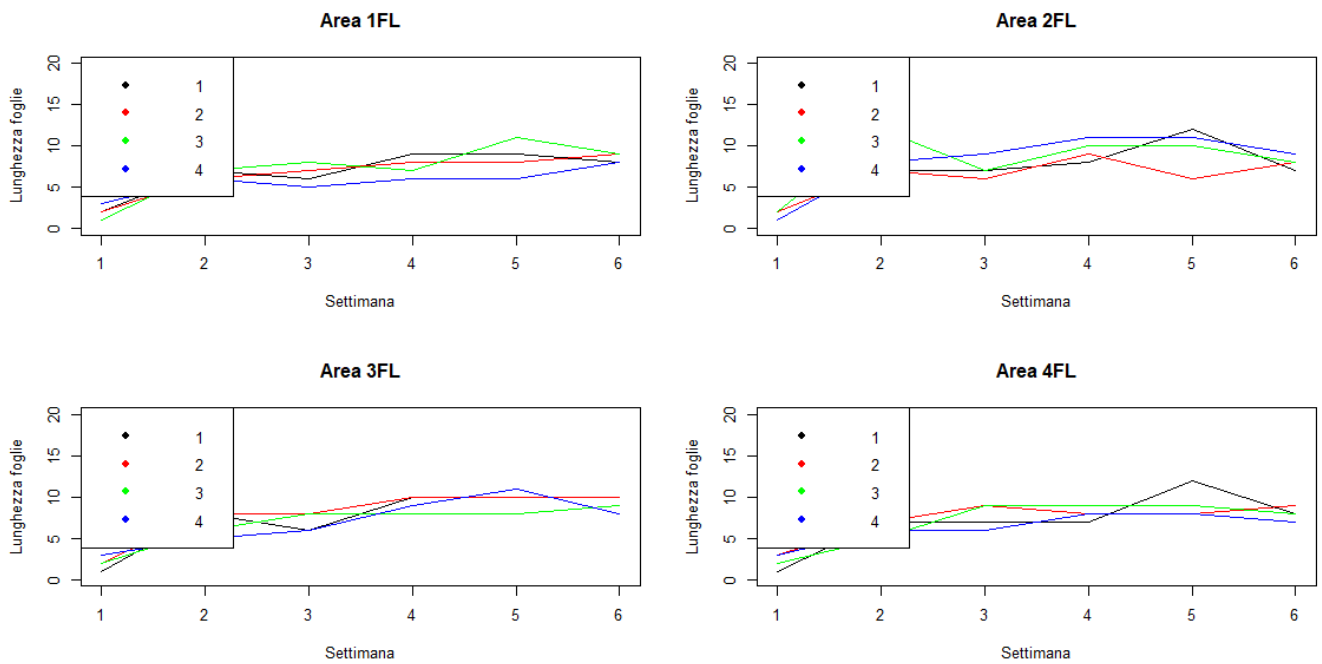


Figura 13. Grafici dell'andamento della lunghezza delle foglie per le patate per aree FL

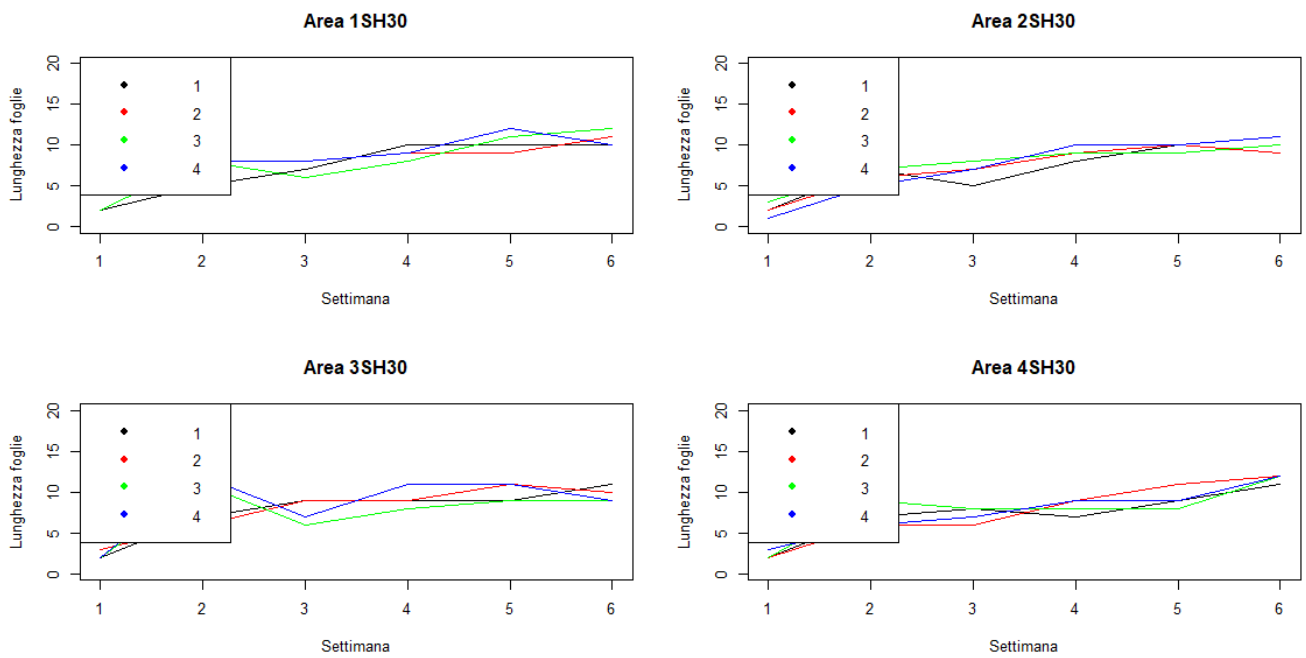


Figura 14. Grafici dell'andamento della lunghezza delle foglie per le patate per aree SH30

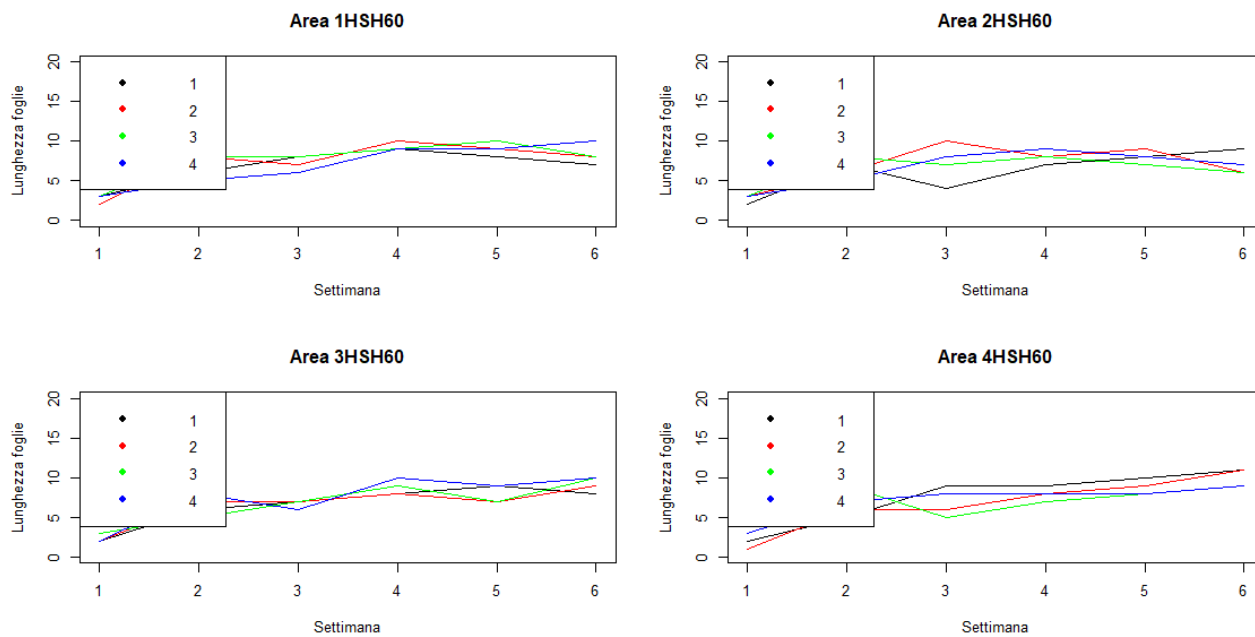


Figura 15. Grafici dell'andamento della lunghezza delle foglie per le patate per aree HSH60

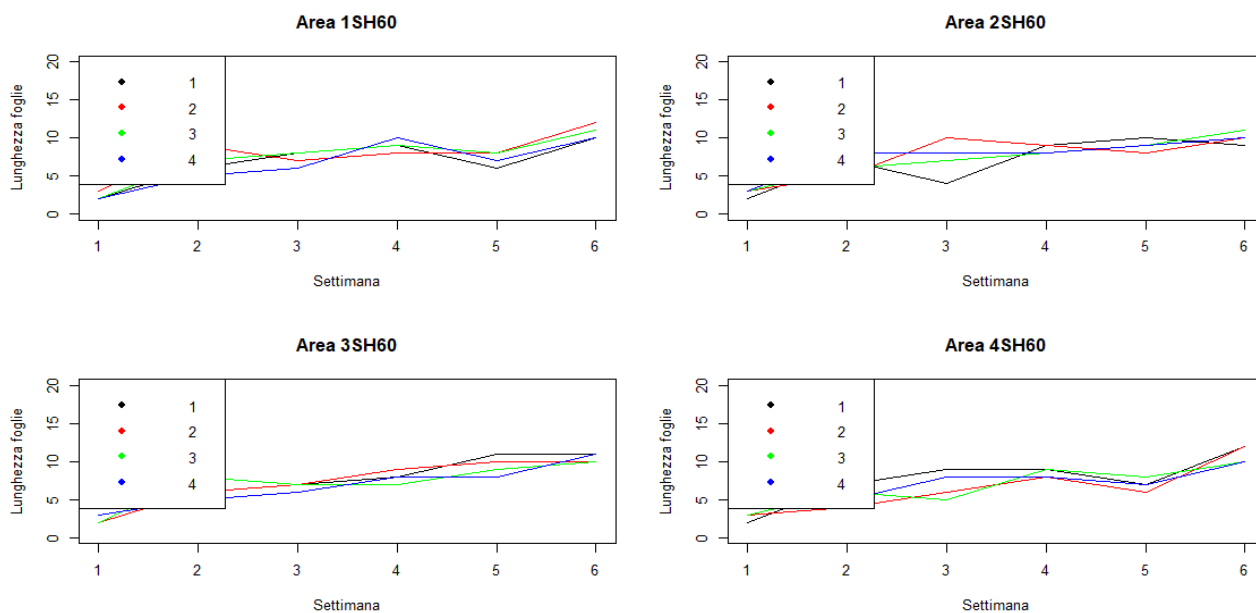


Figura 16. Grafici dell'andamento della lunghezza delle foglie per le patate per aree SH60

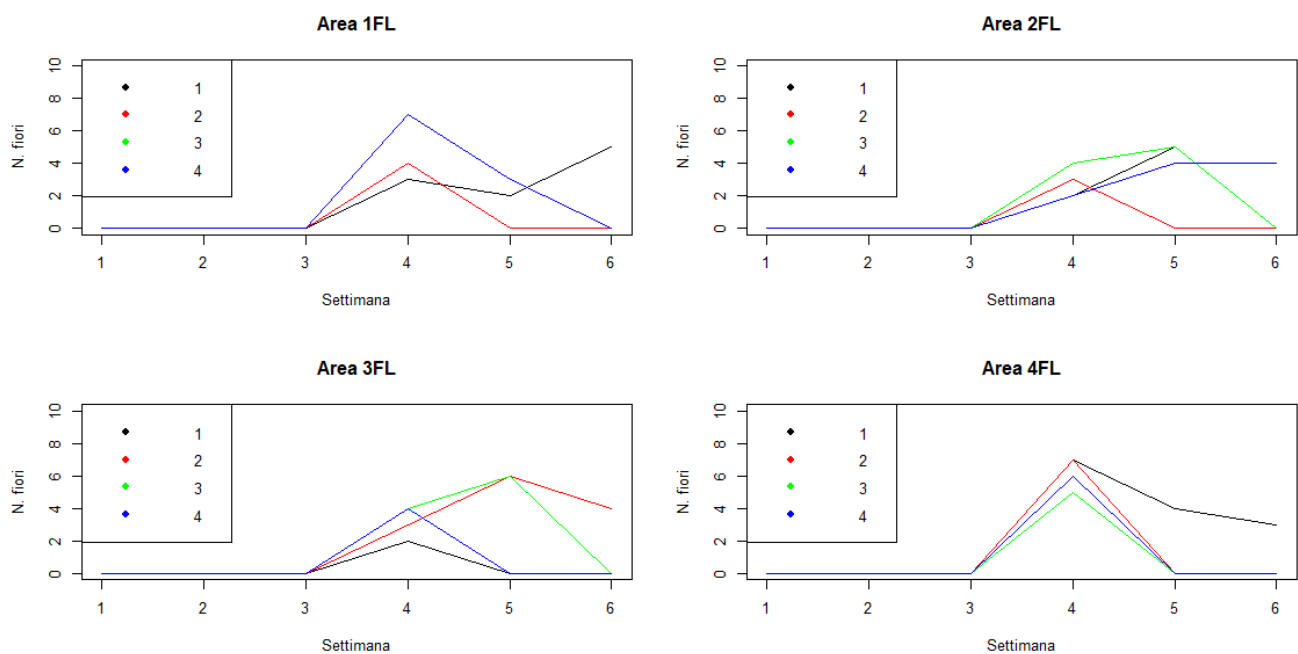


Figura 17. Grafici dell'andamento del numero dei fiori per le patate per aree FL

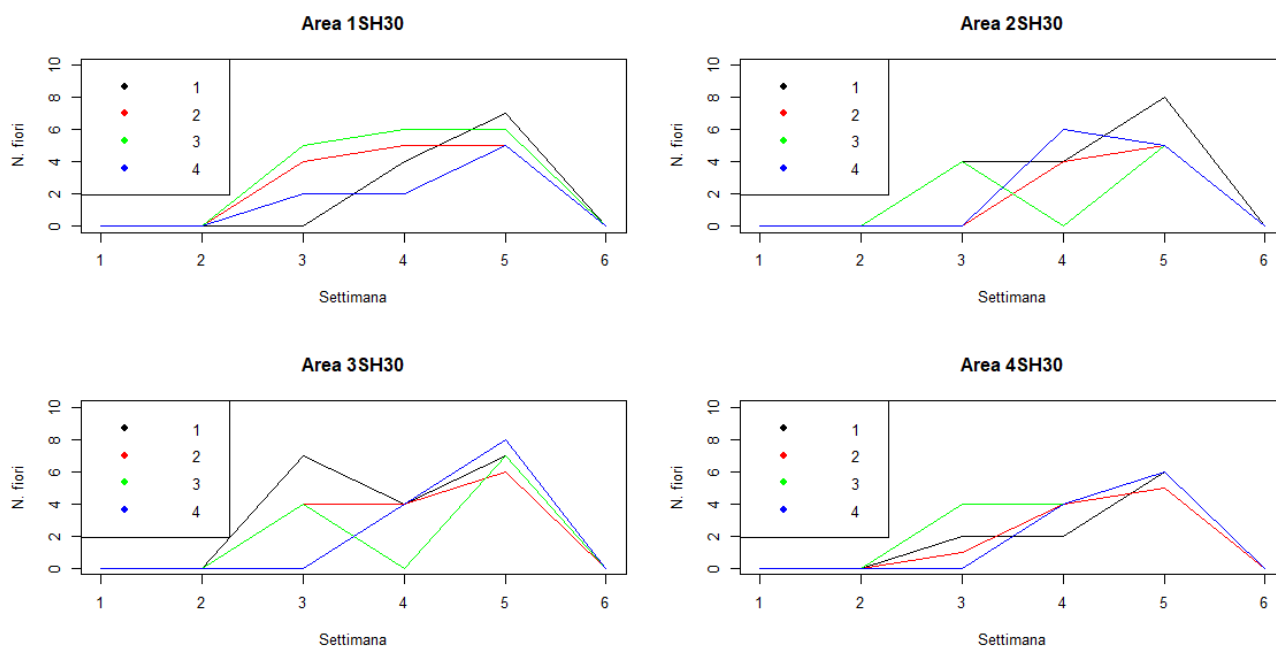


Figura 18.. Grafici dell'andamento del numero dei fiori per le patate per aree SH30

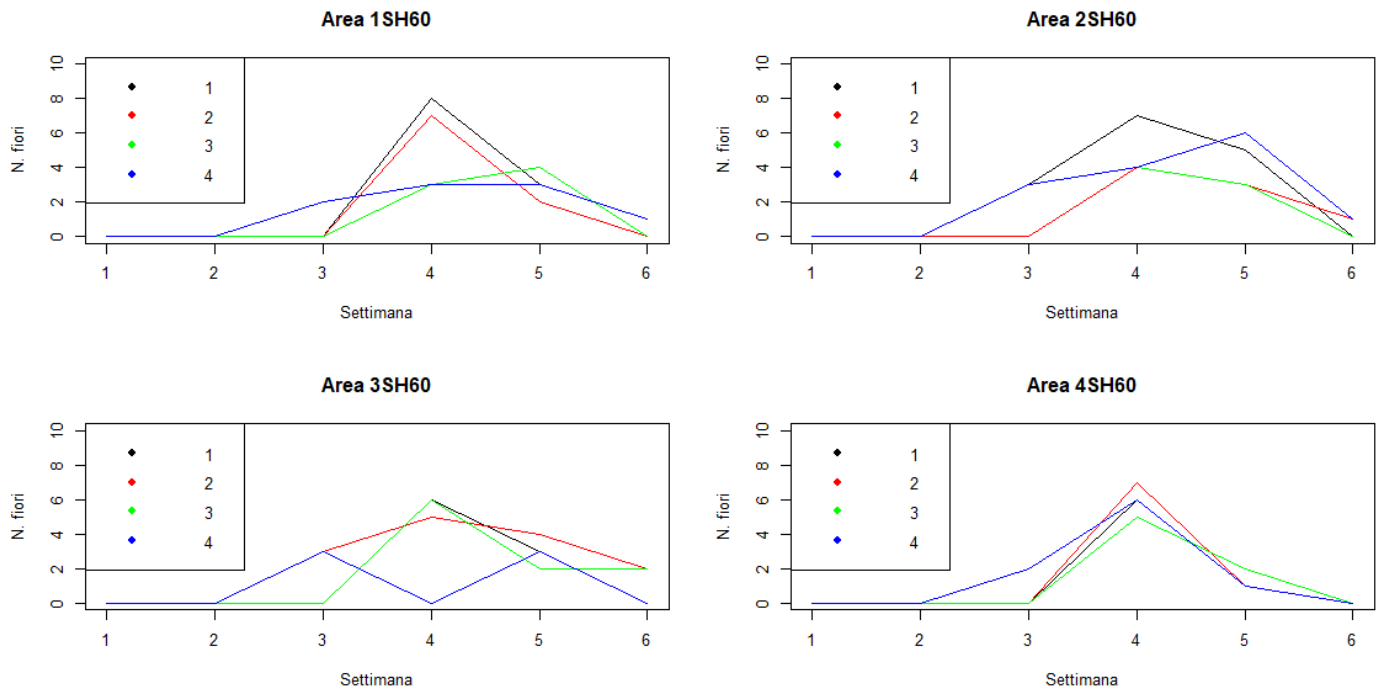


Figura 19.. Grafici dell'andamento del numero dei fiori per le patate per aree SH60

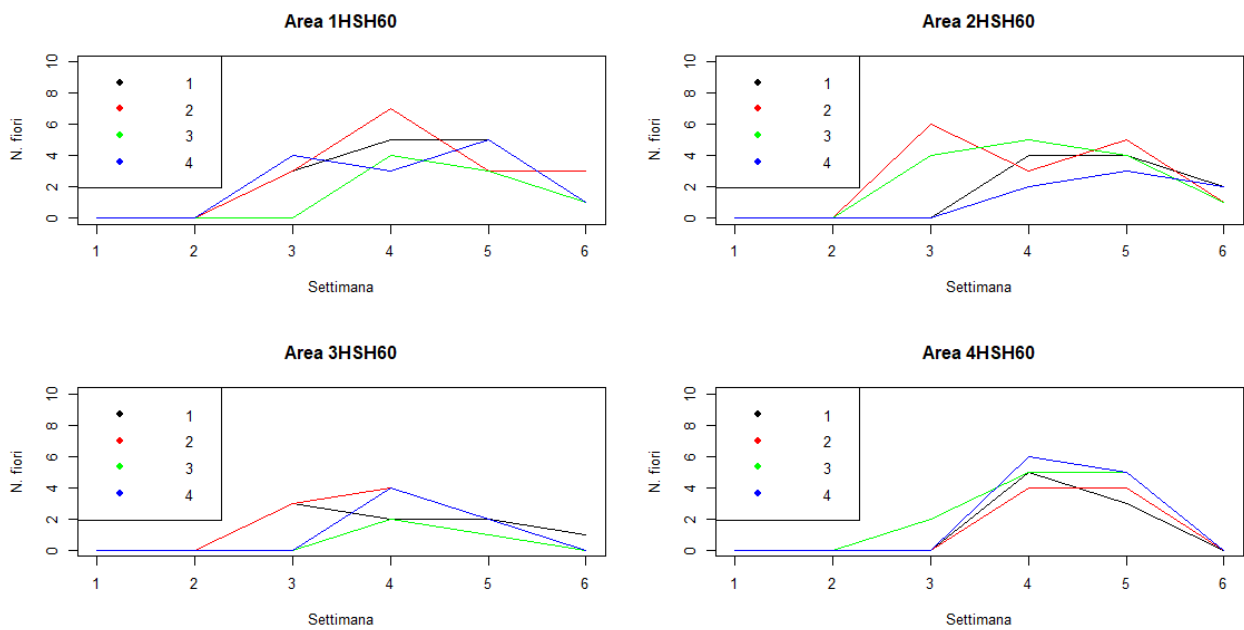


Figura 20. Grafici dell'andamento del numero dei fiori per le patate per aree HSH60