



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA IN  
STATISTICA E TECNOLOGIE INFORMATICHE

TESI DI LAUREA TRIENNALE

**ANALISI DELLA CAPACITA'  
DI DUE CODICI MECCANICI**

Relatore: Ch.ma prof.ssa Giovanna Capizzi

Laureanda: Barbara Corso

Matricola: 483930

ANNO ACCADEMICO 2004/2005



Alla mia famiglia che mi ha sempre sostenuta  
e a Damiano che mi ha sopportata!



# INDICE

	Pag.
• <b>INTRODUZIONE .....</b>	<b>1</b>
• <b>CAPITOLO 1: L'azienda e lo stage.....</b>	<b>3</b>
1.1. L'azienda.....	3
1.2. Lo stage .....	5
1.3. La tesi .....	6
• <b>CAPITOLO 2: I dati .....</b>	<b>7</b>
2.1. La raccolta dei dati .....	7
2.2. Il dataset del Codice 653.....	10
2.3. Il dataset del Codice 755.....	12
• <b>CAPITOLO 3: L'analisi esplorativa dei dati .....</b>	<b>15</b>
3.1. Analisi descrittiva .....	15
3.2. Modelli statistici ed analisi dei residui.....	17
3.3. Codice 653.....	18
3.4. Codice 755.....	26

• <b>CAPITOLO 4: Le carte di controllo.....</b>	<b>37</b>
4.1. Le carte Shewhart.....	38
4.1.1. Misure singole .....	41
4.2 Le carte EWMA.....	42
4.2.1. Misure singole .....	43
4.2.2. Campioni di ampiezza $n>1$ .....	44
4.3. Carte di controllo per dati autocorrelati .....	45
4.4. Applicazione delle carte di controllo al Codice 653 .....	47
4.5. Applicazione delle carte di controllo al Codice 755 .....	51
• <b>CAPITOLO 5: Capacità.....</b>	<b>55</b>
5.1. Indici di capacità.....	55
5.2. Capacità calcolate sul codice 653.....	57
5.3. Capacità calcolate sul codice 755.....	58
• <b>CONCLUSIONI .....</b>	<b>61</b>
• <b>APPENDICE .....</b>	<b>63</b>
A: Grafici .....	63
B: Codice delle funzioni utilizzate.....	78
• <b>RIFERIMENTI BIBLIOGRAFICI.....</b>	<b>83</b>

# INTRODUZIONE

La qualità è diventata una delle principali strategie commerciali e le organizzazioni che si dotano di corretti piani di miglioramento delle qualità riescono ad aumentare la produttività e la penetrazione nel mercato, nonché a raggiungere maggiori profitti e un forte vantaggio competitivo.

Il Controllo Statistico di Processo (SPC *Statistical Process Control*) può essere definito come una metodologia che, in riferimento ad una determinata attività, operazione, fase o processo caratterizzato da ripetitività, fa ricorso a tecniche statistiche al fine di definire, analizzare e verificare le condizioni che determinano la variabilità dell'oggetto di analisi. In altre parole potremmo definire l'SPC come "l'applicazione di tecniche statistiche per comprendere ed analizzare le variabilità di un processo".

Lo scopo dell'analisi consiste nell'ottenere dei dati sui quali fondare determinate conclusioni. Su di essi andranno predisposte azioni per correggere la variabilità del processo in modo tale da risalire alla prima causa su cui bisognerà intervenire per modificare le condizioni di svolgimento del processo stesso.

Nell'individuare i fattori che generano tali anomalie, e quindi variabilità nel processo in esame, si fa sempre riferimento a due categorie principali di cause:

- Cause comuni
- Cause speciali

Le cause comuni (o normali) sono insite nella variabilità di un processo produttivo; insorgono casualmente durante il normale svolgimento del processo e ne determinano la fluttuazione naturale all'interno di un intervallo determinato da un limite di controllo superiore e uno inferiore. Alcuni esempi di cause comuni potrebbero essere: la variazione intrinseca di materiali grezzi utilizzati nella linea produttiva, la mancanza di

un'adeguata supervisione, la vibrazione delle macchine e i cambiamenti nelle condizioni lavorative.

Le cause speciali, invece, sono tutte quelle che, come abbiamo già detto, determinano una variabilità indesiderata o anomala rispetto al naturale svolgimento del processo. Esse possono derivare, ad esempio, dall'uso di un utensile sbagliato, dall'errore di un operatore o da particolari condizioni ambientali, come l'illuminazione o la temperatura. Fin quando non si provvede ad una loro rimozione o correzione esse continueranno ad influire in maniera imprevedibile sul processo portandolo fuori controllo.

Un processo si definisce sotto controllo statistico quando la sua variabilità è dovuta solo a cause comuni. Questa è un aspetto ineliminabile in ogni processo produttivo ma con giusti accorgimenti può essere minimizzata.

Il seguente elaborato ha lo scopo di descrivere l'attività di stage svolta presso l'azienda miniGears collocata nella zona industriale di Camin a Padova.

Nel primo capitolo farò una descrizione dell'azienda nella quale ho svolto la stage ed esporrò l'attività effettuata, mentre nel secondo capitolo descriverò i dati raccolti durante lo stage. Nel terzo analizzerò tali dati mediante grafici e statistiche test e ad essi applicherò dei modelli statistici andando ad analizzarne i residui. Nel quarto capitolo verranno riferite alcune nozioni sulle carte di controllo Shewhart ed EWMA che saranno implementate sui dati. Nel quinto capitolo saranno discussi gli indici di capacità dei processi analizzati.

Seguiranno le conclusioni sul lavoro svolto, un'appendice contenente i grafici dei dati, le funzioni utilizzate nel pacchetto "R" ed una breve bibliografia.



# CAPITOLO 1

## *L'azienda e lo stage*

### 1.1. *L'azienda*

La mG miniGears S.p.A. è stata fondata nel 1976 e fin dall'inizio la sua attività si è concentrata nella produzione di ingranaggi conici e cilindrici in acciaio di piccole dimensioni e di grande serie.



Negli anni successivi l'azienda si è progressivamente sviluppata creando al proprio interno un reparto di sinterizzazione con lo scopo di sviluppare un apposito *know-how* nel particolare settore degli ingranaggi sinterizzati.

Negli ultimi anni mG ha intensificato il proprio processo di internazionalizzazione costituendo un'unità produttiva in USA a Virginia Beach per il mercato nord-americano e un'altra in Cina a Suzhou per il mercato del Far-East: in base a queste iniziative anche le attività di acquisto si sono estese all'estero per cui mG può essere ormai considerata un' "azienda globale".

I settori in cui mG ha operato fino ad oggi con successo sono quelli degli elettroutensili e del giardinaggio: negli ultimi anni mG sta rafforzando la propria presenza nel settore *automotive* con nuove applicazioni rese possibili grazie al proprio particolare *know-how*. La sinergia tra lavorazioni meccaniche e sinterizzazione ha riscontrato molto interesse in diverse applicazioni quali pompe e comandi



pompa, dispositivi di equilibratura motori, sincronizzatori, attuatori per cabriolet, porte scorrevoli, sterzi e freni nonché trasmissioni di potenza con coppie coniche.

La qualità ha rivestito sempre un ruolo molto importante nella politica aziendale di mG e l'azienda è certificata sia ISO/TS 16949 che ISO14001.

Al suo interno essa è divisa in diverse 4 Aree:

- Area 1:
  - lavorazione di dentatura cilindrica;
  - lavorazione di dentatura conica;
  - lavorazione di rettifica;
  - Controllo Qualità Dimensionale (CQD);
  - Uffici acquisti, amministrativo, commerciale, direzione, personale, magazzino.
- Area 2:
  - sinterizzazione (compattazione e trattamento termico del sinterizzato);
  - reparto di trattamento termico dell'acciaio;
  - Controllo Qualità Metallurgico (CQM);
  - Controllo Qualità Sinterizzato (CQS).
- Area 3:
  - reparto di manutenzione;
  - rettifica frese (utensili per la lavorazione);
  - CQD per collaudo lavorazioni in produzione;
  - tornitura.
- Area 9:
  - montaggio testine (decespugliatori);
  - montaggio corpi di trasmissione per muletti;
  - CQD;
  - lavorazione meccaniche.

## 1.2. *Lo stage*

L'esperienza di stage presso la mG è durata 3 mesi, nel corso dei quali ho preso dimestichezza con strumenti per il controllo dimensionale, software statistici e, per la prima volta sono stata immersa in una grande realtà aziendale.

Il primo periodo è stato dedicato ad un orientamento nell'azienda, alla conoscenza delle persone che vi lavorano ed alla sistemazione nel settore del Controllo Qualità Dimensionale (*Control Quality Dimensional CQD*) collocato nell'Area 1.

L'azienda era interessata ad una mappatura del reparto rettifiche, ovvero ad un'analisi della capacità delle macchine del reparto nel quale vengono "portati a misura" i fori e i diametri dei vari ingranaggi.

Il reparto rettifiche è costituito di circa 20 macchine ognuna delle quali lavora un ingranaggio diverso, ad eccezione di due che lavorano lo stesso codice in quanto richiesto in quantità elevate.

Le misure sono state effettuate direttamente nel reparto per non incappare in una differenza di rilevazione dovuta allo sbalzo termico tra settore di rettifica e CQD.

Gli operatori effettuano un campionamento basato su 3 pezzi l'ora, secondo le normative. Tutte le rilevazioni vengono inserite nel computer ed analizzate mediante istogrammi e carte di controllo. Considerando queste analisi si è ipotizzata la presenza di una instabilità del processo osservato. Da questa valutazione ho pensato di estrarre campioni di 50 osservazioni consecutive allo scopo di esplorare la presenza di una correlazione seriale e di eventuali salti significativi nella media del processo.

### 1.3. *La tesi*

Tra tutti i codici analizzati durante lo stage, per lo svolgimento della tesi si è scelto di studiarne solamente 2.

Il primo codice, che chiameremo Codice 653 per la parte finale del suo *part number*, è stato scelto in quanto facente parte di una serie di codici *automotive*.

Il secondo, Codice 755, è stato individuato in quanto prodotto da un macchinario piuttosto vecchio e, per questo motivo, avente dei problemi.

Su questi due codici sono state fatte delle misurazioni sui fori, da tali misure si è proceduto con delle analisi esplorative, con l'adattamento di convenienti modelli statistici ed infine con l'applicazione della carte di controllo. Inoltre, per valutare il rendimento del processo produttivo, sono stati calcolati alcuni indici di capacità.

Per i grafici e per le analisi dei dati e dei modelli inseriti nell'elaborato, sono stati usati il software "R" e "MiniTab". Nell'Appendice sono riportati alcuni grafici, che non sono stati introdotti nella stesura e nell'analisi della relazione, e le funzioni utilizzate per i disegni delle carte di controllo.

## CAPITOLO 2

### *I dati*

#### *2.1. La raccolta dei dati*

Come detto i dati si riferiscono a dei fori di ingranaggi meccanici e le misurazioni sono state effettuate nell'istante in cui il pezzo usciva dalla lavorazione. Le rilevazioni sono state eseguite a lato delle macchine dove vi è una temperatura pressoché costante all'incirca pari a 23°. Operando in questo modo si può assumere che i dati siano tutti posti nelle stesse condizioni atmosferiche.

I pezzi entrano in macchina, vengono lavorati mediante una mola e contemporaneamente raffreddati per mezzo di getti di acqua emulsionata (un mix di acqua ed olio) ed escono dalla macchina grazie ad un nastro trasportatore. Da qui vengono oliati per evitare la ruggine, messi su delle griglie a sgocciolare e poi sistemati in cassette che successivamente verranno mandate al magazzino spedizioni nell'Area 1 dove i pezzi saranno impacchettati, imballati ed inviati ai clienti.

Le macchine che producono i due codici qui analizzati vengono accese dall'operatore che effettua il turno della mattina (6 – 13:30) e spente durante la notte dal turnista della sera (13:30 - 21). Abbiamo quindi un blocco della macchina e di conseguenza un periodo iniziale di assestamento una volta riaccesa.

Per le rilevazioni il pezzo è stato prelevato una volta uscito dalla macchina, pulito dai residui di detergente con delle pistole ad aria compressa o con dei panni, misurato e depositato nelle cassette da portare al magazzino.

Per le misure sono stati utilizzati dei micrometri da interni con tolleranza  $10 \div 12$  per il Codice 653 e tolleranza  $6 \div 8$  per il Codice 755.

I micrometri venivano collegati al computer mediante un cavo così da poter inserire le misure direttamente in un foglio di calcolo, nel nostro caso Excel. ( Figura 2.1. )

<b>mG</b>				Data	
miniGears					
<b>DATI RACCOLTI SU p/n _____ CDL. _____</b>					
<b>QUOTA <math>\varnothing</math> _____ TOLLERANZA _____</b>					
n° 1			n° 26		
n° 2			n° 27		
n° 3			n° 28		
n° 4			n° 29		
n° 5			n° 30		
n° 6			n° 31		
n° 7			n° 32		
n° 8			n° 33		
n° 9			n° 34		
n° 10			n° 35		
n° 11			n° 36		
n° 12			n° 37		
n° 13			n° 38		
n° 14			n° 39		
n° 15			n° 40		
n° 16			n° 41		
n° 17			n° 42		
n° 18			n° 43		
n° 19			n° 44		
n° 20			n° 45		
n° 21			n° 46		
n° 22			n° 47		
n° 23			n° 48		
n° 24			n° 49		
n° 25			n° 50		

Figura 2.1: Scheda raccolta dati

Bisogna tener presente che ci sono due cose importanti da eseguire prima delle rilevazioni.

Una da effettuare sempre è quella di azzerare lo strumento di misura. Ciò va fatto con degli appositi blocchetti in acciaio o ceramica per calibri o

micrometri da esterni e con degli anelli per i micrometri da interni. I blocchetti e gli anelli sono disponibili in varie dimensioni ed ogni strumento ha la sua taratura. Questa operazione consiste nel togliere con dell'alcool eventuali residui di olio o di trucioli dalle testine misuratrici dello strumento e nel pulire anche il blocchetto che in seguito verrà misurato. Se la lettura da strumento non corrisponde con la grandezza del blocchetto, si deve reimpostare la taratura sullo strumento. Ciò è importante in quanto si potrebbero eseguire delle rilevazioni che sembrano corrette ed invece non sono affidabili.

L'altra cosa da effettuare frequentemente è una prova di Ripetibilità e Riproducibilità degli strumenti. Questa operazione viene eseguita da 3 o più operatori i quali prendono 10 pezzi, li numerano e a turno li misurano prendendoli in ordine casuale, tenendo però nota del numero identificativo del pezzo misurato. Tutte queste azioni andranno ripetute 3 volte per ciascun operatore; il fatto di aver annotato le misurazioni relative al primo pezzo consente di poter calcolare l'errore commesso da ciascun operatore, quindi la ripetibilità.

Le misure possono cambiare oltre che per problemi dovuti allo strumento anche da operatore ad operatore. Per questo i dati utilizzati per la tesi e per la mappatura del reparto rettifiche si riferiscono tutti a misure rilevate personalmente dalla sottoscritta.

Prima di poter ritenere attendibili le mie rilevazioni ho dovuto prendere un po' di manualità con gli strumenti di misura. Ho appreso la giusta posizione del corpo rispetto allo strumento, il punto di posizionamento del pezzo sottoposto a misurazione rispetto ai punti di contatto dello strumento e la giusta pressione da applicare alla frizione del micrometro.

## 2.2. Il dataset del Codice 653

Il Codice 653 fa riferimento ad un ingranaggio cilindrico con 37 denti ed è costituito da materiale sinterizzato (Figura 2.2).



Figura 2.2: Ingranaggio cilindrico, Codice 653

Il sinterizzato consiste in un particolare tipo di polvere che viene compattata con degli appositi stampi in maniera tale da creare la forma desiderata. Questo prodotto viene successivamente scaldato in dei forni per far legare tra loro le molecole contenute nella polvere ed infine viene lavorato secondo le esigenze.

Le rilevazioni fanno riferimento ad un arco di tempo che va dal 23 marzo al 13 aprile 2005 e sono state effettuate tutte durante il turno di uno stesso operatore. Questo codice rappresenta uno dei particolari *automotive* costruiti da mG, ciò significa che è sottoposto a dei regimi di qualità molto più rigidi. Si esige quindi una buona resa da parte della macchina che lo produce. Questa ha al suo interno dei tastatori i quali vanno a verificare la quota subito dopo la lavorazione in modo da avere un controllo al 100% ed affidabile. Ciò fa sì che l'operatore non debba effettuare particolari interventi di correzione sulle misure ma sia tenuto "solo" a controllare i 3 pezzi l'ora, a caricare i pezzi in macchina affinché vengano lavorati e porli nelle rastrelliere a fine rettifica.



Le misure fanno riferimento al foro dell'ingranaggio 653 che ha come valore di target 10.026 mm ed un campo di tolleranza di  $\pm 0.008$  mm avendo quindi un range che va da 10.018 mm a 10.034 mm.

Come si vede in Figura 2.3 il dataset è costituito da 300 osservazioni rilevate in 6 giorni diversi. Abbiamo quindi la presenza di 6 gruppi da 50 dati ciascuno che identificheremo, per motivi di semplicità, con le lettere dell'alfabeto: A, B, C, D, E ed F.

### Misurazioni foro Codice 653

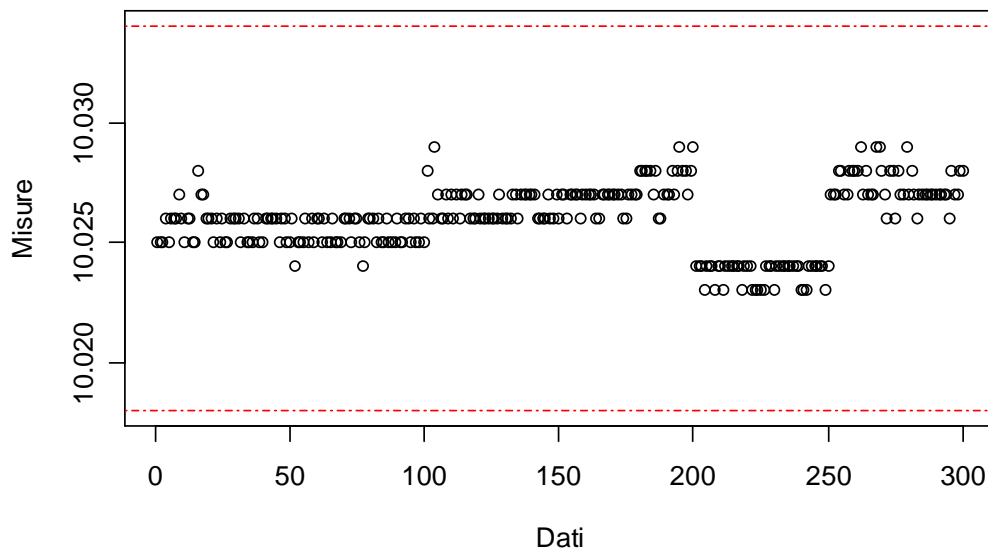


Figura 2.3: Dataset Codice 653

Guardando i dati si può osservare come essi restino ampiamente all'interno dei limiti di tolleranza evidenziati dalle linee rosse tratteggiate. Si riscontra un andamento piuttosto costante della media all'incirca fino all'osservazione 200, da tale momento in poi si nota un abbassamento del livello medio del processo che torna a crescere dopo la 250° osservazione (tali valori appartengono al gruppo identificato con la lettera E, vedremo nel Capitolo 3.1. le analisi adottate per questi dati).

### 2.3. *Il dataset del Codice 755*

Il Codice 755 è un pignone conico a denti dritti ed anch'esso è costituito da materiale sinterizzato (Figura 2.4).



Figura 2.4: Pignone conico a denti dritti, Codice 755

Le rilevazioni sono state eseguite in un arco di tempo che va dal 16 marzo al 13 aprile 2005 e questa volta non fanno riferimento ad uno stesso operatore, ma a 4 addetti diversi.

Inoltre la macchina non possiede, come visto per il Codice 653, un sistema di rilevazione automatico al suo interno. In questo modo l'operatore sarà costretto ad effettuare un controllo al 100% mediante un tampone passa – non passa e se si verificano delle variazioni sulle misure dovrà attuare degli interventi correttivi.

Le misure fanno riferimento al foro dell'ingranaggio 755 che ha come valore di target 7 mm ed un campo di tolleranza di + 0.012 / + 0.022 mm, con un range che va da 7.012 mm a 7.022 mm.

Come si vede in Figura 2.5 il dataset è costituito da 400 osservazioni rilevate in 8 giorni diversi. Anche in questo contesto abbiamo dei gruppi che identificheremo sempre con le lettere dell'alfabeto: A, B, C, D, E, F, G ed H che contengono 50 osservazioni ciascuno.

## Misurazioni foro Codice 755

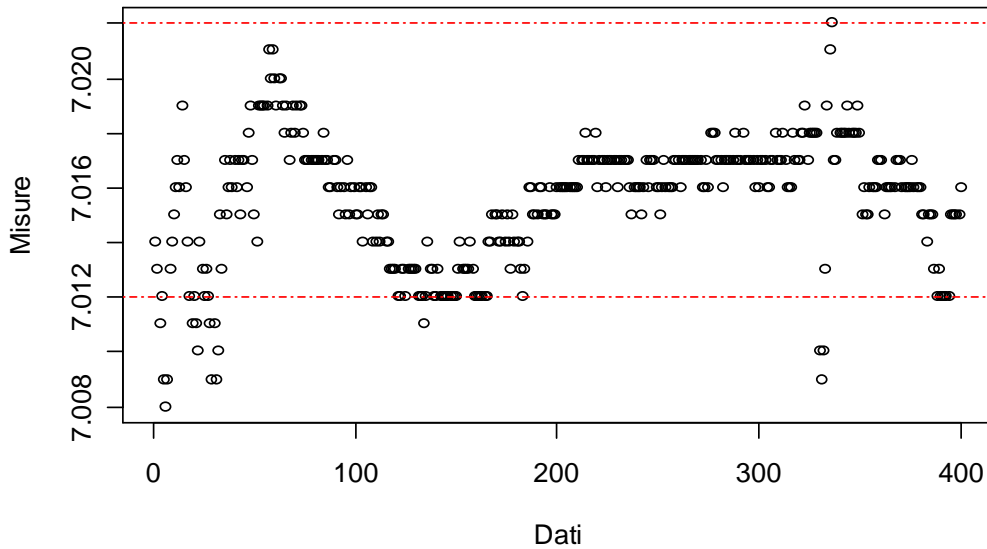


Figura 2.5: Dataset Codice 75

Diversamente dal caso precedente, i dati non presentano un andamento stabile nel tempo e si può notare che in essi è identificabile un certo trend a tratti. Le osservazioni iniziali sono crescenti fino quasi alla 70°, abbiamo poi una ripida decrescita fino a 150 circa seguita da una lenta risalita fino all'osservazione 320. Nella parte finale si riscontra una leggera diminuzione nei valori rilevati.

Nel grafico sono stati tratteggiati con delle linee rosse i limiti di tolleranza del diametro della variabile in esame. Si può vedere come molte osservazioni cadano all'esterno di tali limiti con un'elevata variabilità.



## CAPITOLO 3

### *L'analisi esplorativa dei dati*

#### 3.1 *L'analisi descrittiva*

I dataset dei due codici contengono misurazioni su di una caratteristica dell'output di due processi produttivi, esattamente il diametro (in mm) del foro di due ingranaggi meccanici. I valori registrati sono invece le deviazioni dei diametri dal valore obiettivo (target). Per l'azienda produttrice l'ideale sarebbe ridurre il più possibile le fluttuazioni intorno ai rispettivi valori target, ovvero riuscire ad avere degli output il più possibile stabili nel tempo e vicini ai valori obiettivo.

Tutte le misure presenti nei dataset sono state rilevate in maniera sequenziale. Durante l'analisi dovremo quindi tener conto della elevata autocorrelazione all'interno dei gruppi.

Sulla base di tali dati si vuole capire se le oscillazioni osservate sono puramente casuali o se piuttosto presentano una qualche forma di dipendenza dal tempo oppure da altri possibili fattori.

Per verificare la normalità costruiremo degli istogrammi e dei boxplot, tuttavia per un'analisi più approfondita utilizzeremo il *test di Shapiro-Wilk*, dal nome degli autori che lo hanno introdotto.

Il test è basato sul *normal probability plot* e vuol saggiare le seguenti ipotesi:

$$\begin{cases} H_0: \text{la distribuzione dei dati è normale} \\ H_1: \text{la distribuzione dei dati non è normale} \end{cases}$$

La statistica test è sostanzialmente il coefficiente di correlazione tra le coppie di punti riportate sul *normal probability plot*.

Per verificare che il campione di dati provenga da una distribuzione normale di media  $\mu_0$  prestabilita si considera il sistema di verifica d'ipotesi:

$$\begin{cases} H_0: \hat{\mu} = \mu_0 \\ H_1: \hat{\mu} \neq \mu_0 \end{cases}$$

Sotto l'ipotesi nulla i dati si distribuiscono come una variabile casuale Normale di media  $\mu_0$  e varianza non nota  $\frac{\sigma^2}{n}$ . Tale varianza andrà stimata mediante la varianza campionaria  $s^2$  approssimando così la distribuzione con una *t di Student* con  $n - 1$  gradi di libertà.

La statistica test risulta pari a:

$$t^{oss} = \frac{\hat{\mu} - \mu_0}{s} \cdot \sqrt{n}$$

dove con  $\hat{\mu}$ ,  $s$  e  $n$  indichiamo rispettivamente la media, la deviazione standard campionaria e la numerosità dei campioni.

Con un *test t a due campioni* si verifica se le medie di due gruppi risultano significativamente diverse. Le ipotesi del test sono:

$$\begin{cases} H_0: \mu_A = \mu_B \\ H_1: \mu_A \neq \mu_B \end{cases}$$

Per procedere in questa analisi dobbiamo calcolare una nuova varianza:

$$s^2 = \frac{1}{n + m - 1} \left[ \sum_{i=1}^n (y_{ii} - \bar{y}) + \sum_{i=1}^m (x_i - \bar{x}) \right]$$

dove  $n$  ed  $m$  rappresentano le numerosità campionarie ed  $\mu_A$  e  $\mu_B$  le rispettive medie campionarie. Il test risulta pari a:

$$t^{oss} = \frac{\bar{y} - \bar{x}}{s} \cdot \sqrt{\frac{1}{n} + \frac{1}{m}}$$

Sotto l'ipotesi nulla la statistica test si distribuirà come una *t di Student* con  $n + m - 2$  gradi di libertà.

Per verificare se i dati sono autocorrelati andremo ad applicare dei diagrammi di correlazione totale e parziale. Infine presenteremo dei grafici di autodispersione per vedere fino a che ritardo le variabili sono legate.

### 3.2 Modelli statistici ed analisi dei residui

Per lo studio della relazione tra due variabili per prima cosa viene proposta la retta di regressione lineare semplice:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad 3.1$$

dove  $y$  è la variabile risposta,  $\beta_0$  l'intercetta,  $\beta_1$  il coefficiente della esplicativa  $x$  ed  $\varepsilon$  un termine non osservabile di errore casuale.

Tale errore è supposto di media nulla e varianza costante ma ignota  $\sigma^2$ .

Una notazione più generale della 3.1 è data da:

$$y = f(x; \beta) + \varepsilon \quad 3.2$$

che dà luogo alla 3.1 quando  $f$  assume l'espressione di una retta.

Solitamente le stime dei coefficienti  $\beta$  avvengono tramite il criterio dei minimi quadrati e si trova:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad 3.3$$

dove  $X$  è la matrice di regressione contenente i valori delle variabili esplicative.

Per valutare il grado di adattamento del modello si calcola il coefficiente di determinazione:

$$R^2 = 1 - \frac{\text{devianza residua}}{\text{devianza totale}} \quad 3.4$$

Per una verifica della bontà di adattamento è necessario completare lo studio effettuando un controllo empirico mediante l'analisi dei residui

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad 3.5$$

Uno strumento molto utile in questo caso è dato dall'analisi grafica. Si costruiscono:

- istogrammi e *normal probability plot* per individuare eventuali deviazioni dall'ipotesi di normalità degli errori;
- grafico dei residui contro i valori stimati per visualizzare carenze sistematiche dei modelli (evidenziate da andamenti particolari dei residui in funzione dei valori stimati), riscontrare violazioni dell'ipotesi di omoschedasticità ed individuare osservazioni mal colte dal modello;
- grafico dei residui standardizzati contro i valori stimati per cogliere la presenza di eventuali valori anomali

Prima di scegliere i modelli utilizzati nell'analisi, è stato effettuato uno studio sulle variabili. Operando delle trasformazioni abbiamo verificato se vi fossero o meno dei legami particolari tra di esse ma non abbiamo trovato delle trasformazioni apprezzabili.

### 3.3 Codice 653

Il dataset del Codice 653 contiene 300 misurazioni suddivise in 6 campioni di ampiezza 50 ciascuno ed è costituito dalle seguenti variabili:

- 'Foro' : variabile continua contenente le misure dei diametri;
- 'Gruppo' : variabile categoriale a 6 livelli indicante il gruppo di appartenenza;
- 'Trend' : variabile intera che tiene conto dell'effetto temporale.

Il valore target del processo è 10.026 mm e come si vede dal grafico in Figura 3.1 i campioni risultano piuttosto diversi in media:



### Carta EWMA per le misure singole, lambda = 0.45

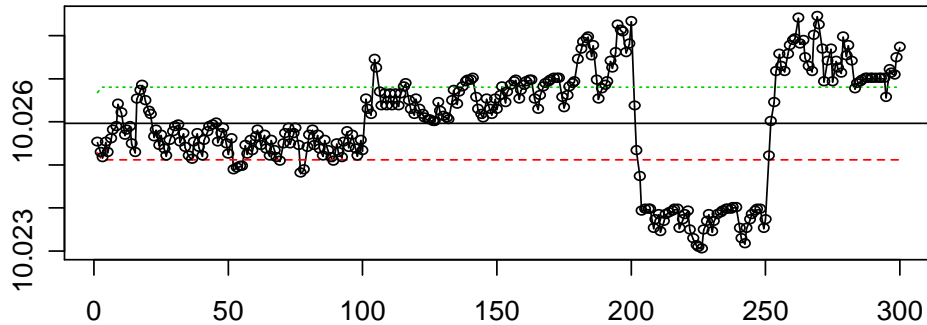


Figura 3.1: Carta EWMA per le misure singole del Codice 653 con  $\lambda$  pari a quello ottimale.

La differenza nel parametro di posizione è chiaramente visibile dai boxplot dei dati per Gruppo (Figura 3.2):

### Misurazioni nei gruppi, Codice 653

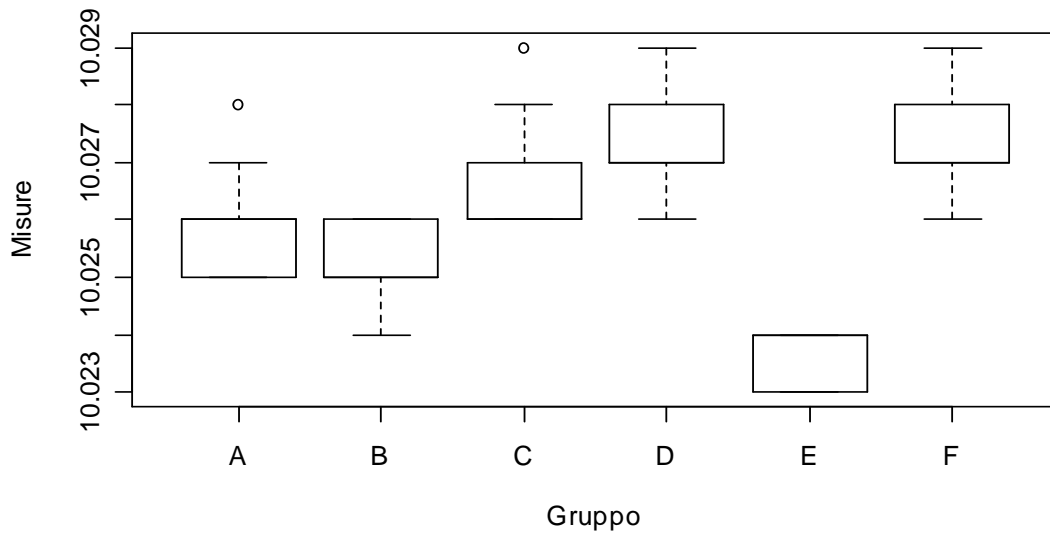


Figura 3.2: Boxplot dei dati del Codice 653

In particolare la media complessiva dei dati è vicina al valore di target previsto dal processo. Le medie dei Gruppi A e B sono molto vicine tra loro, così pure quelle relative ai Gruppi C e D. Si può pensare di

suddividere l'insieme di dati in due nuovi gruppi: uno contenente le osservazioni dei campioni A e B ed uno contenente quelle dei campioni C e D. Il gruppo E risulta quello con il più basso livello di processo, mentre al contrario il Gruppo F assume il valor medio maggiore.

Verifichiamo se i dati si scostano o meno dal valore nominale calcolando la statistica test *t di Student*. Otteniamo le seguenti statistiche affiancate ai relativi livelli di significatività osservati:

	$t^{oss}$	$\alpha^{oss}$
<b>Totale</b>	- 0.3741	0.7086
<b>Gruppo A</b>	- 3.1305	0.00294
<b>Gruppo B</b>	- 7.4246	1.463e-09
<b>Gruppo C</b>	5.25	3.265e-06
<b>Gruppo D</b>	11.062	6.352e-15
<b>Gruppo E</b>	- 35.5457	<2.2e-16
<b>Gruppo F</b>	12.9597	<2.2e-16

Come si può facilmente vedere dai livelli di significatività associati a ciascun test solamente le media complessive dei dati porta all'accettazione dell'ipotesi nulla. Quindi nel complesso il processo non si scosta dal valore di target prefissato, ma le medie dei singoli campioni si allontanano significativamente dal valore obiettivo.

Dato che dai boxplot abbiamo detto che i gruppi A e B e i gruppi C e D hanno medie molto vicine proviamo ad applicare un *test t a due campioni*.

I valori che otteniamo sono:

$$\begin{aligned}
 t_{AB}^{oss} &= 2.3932 & \alpha_{AB}^{oss} &= 0.01861 \\
 t_{CD}^{oss} &= -4.7907 & \alpha_{CD}^{oss} &= 5.908e-06
 \end{aligned}$$

entrambi ci portano quindi a rifiutare l'ipotesi nulla.

La nostra intenzione di raggruppare questi due gruppi non potrà quindi essere compiuta.

Andiamo ora ad analizzare la normalità dei dati, mediante un istogramma con la relativa funzione di densità stimata ed un boxplot (Figura 3.3):

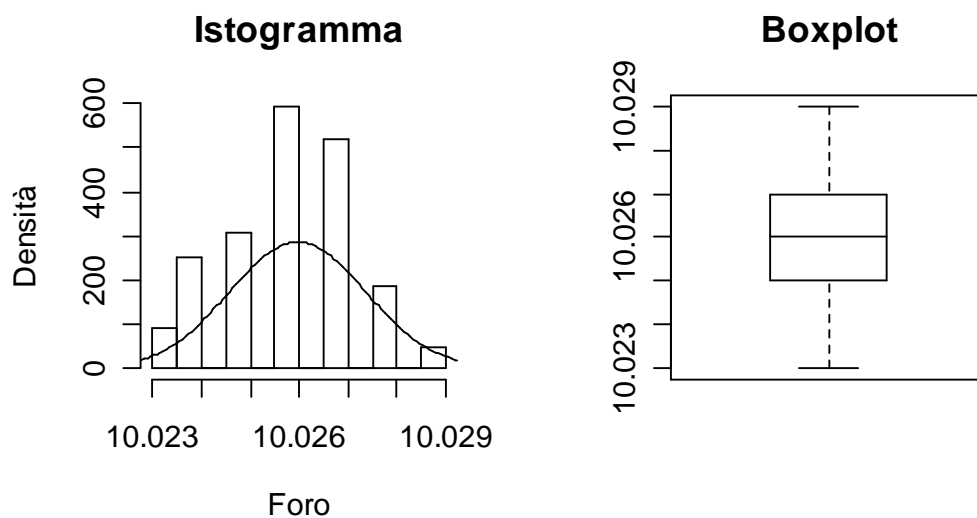


Figura 3.3: Istogramma con densità e boxplot dei dati del Codice 653

Entrambi i grafici fanno pensare ad un andamento normale: la distribuzione dei dati è simmetrica, e ciò è ben visibile dalla scatola del boxplot e dalla lunghezza dei baffi, non vi sono valori anomali e nemmeno dei salti nelle barre dell'istogramma. Tuttavia applicando il *test di Shapiro-Wilk* tale ipotesi è completamente violata.

Il *normal probability plot* nel caso in esame è rappresentato in Figura 3.4:

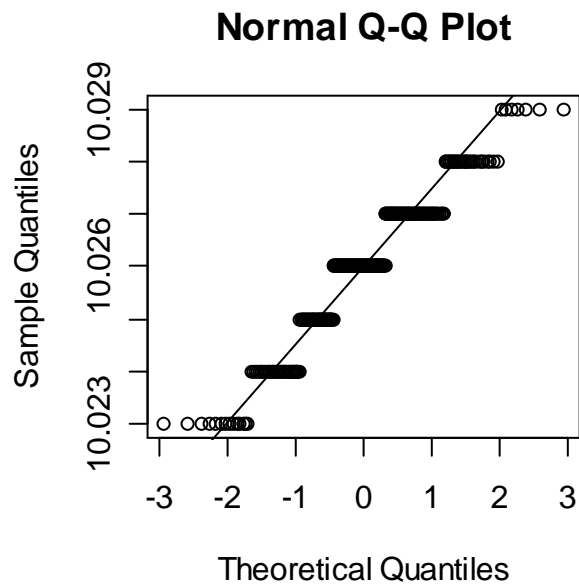


Figura 3.4.: Normal probability plot dei dati del Codice 653

ed il valore della statistica test per i nostri dati risulta:

$$W = 0.9414$$

con un livello di significatività osservato pari a:

$$\alpha^{oss} = 1.562e - 09$$

questi valori ci portano a rifiutare l'ipotesi di normalità per la distribuzione dei dati.

Per avere un'idea della correlazione seriale riportiamo in Figura 3.5 un grafico sulla correlazione totale ed uno sulla parziale:

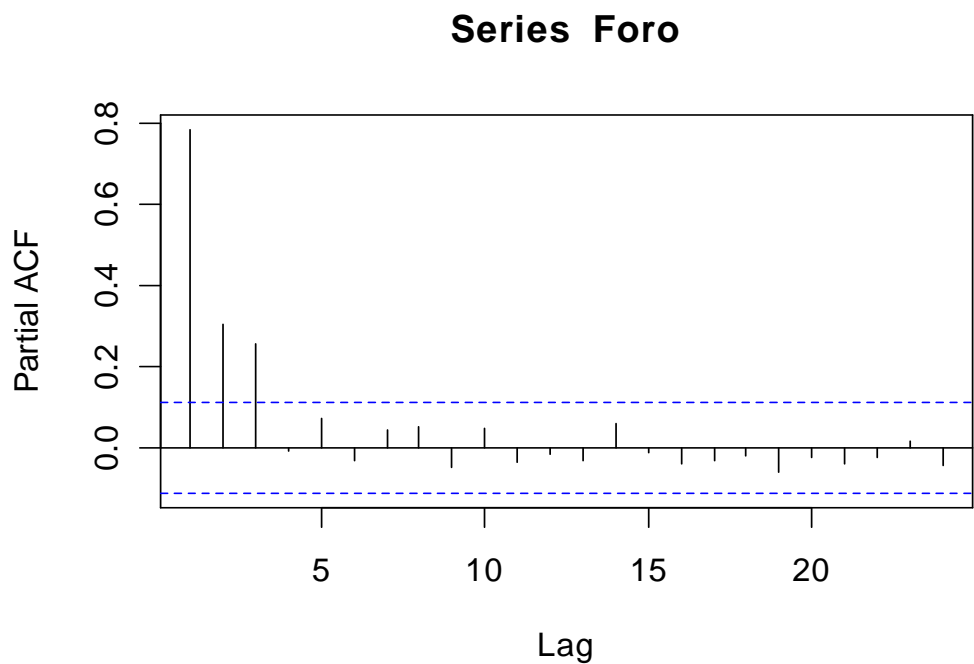
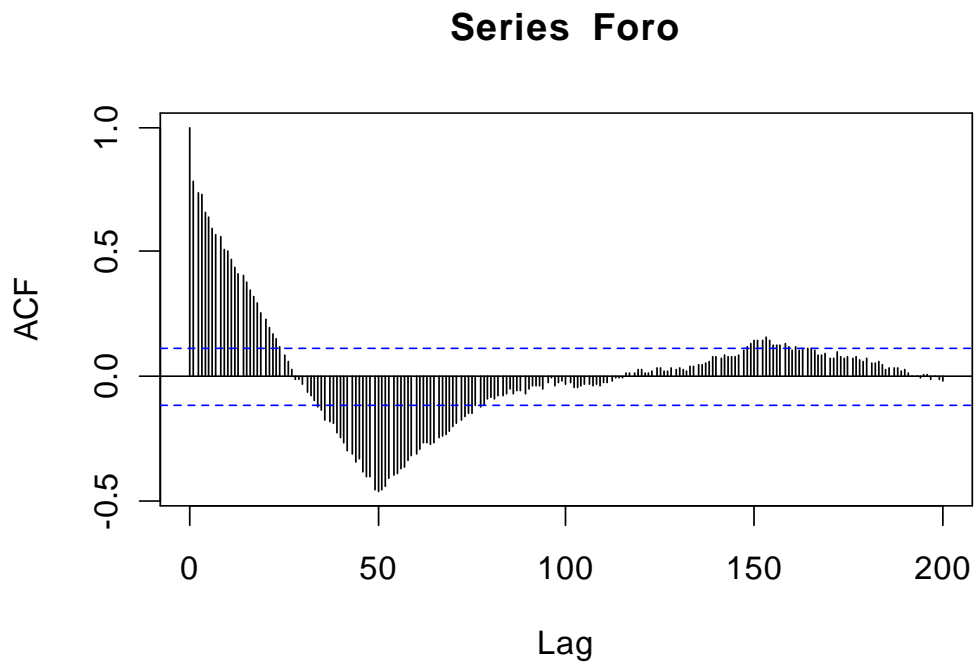


Figura 3.5: Grafico della autocorrelazione totale e parziale, Codice 653

Dal grafico totale si può affermare che c'è una forte autocorrelazione sia positiva sia negativa. Dal grafico parziale i primi tre coefficienti sembrano

significativamente differenti da zero (escono dalle bande), tutti gli altri no. Questo suggerisce come possibile modello per interpolare i dati un MA(3).

Dunque il nostro dataset è costituito da 300 osservazioni su 3 variabili. Tali osservazioni non seguono un andamento normale e sono fortemente autocorrelate.

Per quanto riguarda i modelli adottati, nel dataset abbiamo prima utilizzato un modello avente come risposta la variabile 'Foro' e come esplicative 'Trend' e 'Gruppo'. Il livello di significatività associato all'effetto 'Trend' non risultava significativo, perciò tale variabile è stata tolta dal modello. Una sintesi del modello adattato è riportata di seguito:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.003e+01	9.142e-05	109665.019	< 2e-16	***
GruppoB	-3.000e-04	1.293e-04	-2.320	0.021	*
GruppoC	7.800e-04	1.293e-04	6.033	4.83e-09	***
GruppoD	1.440e-03	1.293e-04	11.138	< 2e-16	***
GruppoE	-1.980e-03	1.293e-04	-15.315	< 2e-16	***
GruppoF	1.680e-03	1.293e-04	12.994	< 2e-16	***

Residual standard error: 0.0006464 on 294 degrees of freedom  
 Multiple R-Squared: 0.787, Adjusted R-squared: 0.7833  
 F-statistic: 217.2 on 5 and 294 DF, p-value: < 2.2e-16

L'output ci fornisce una stima dei coefficienti affiancati dal corrispondente standard error, il valore della statistica *t di Student* ed il p-value associato. Come si può notare i coefficienti sono tutti significativi per spiegare il modello.

L' $R^2$  ci dice che il modello adattato spiega in modo soddisfacente la variabilità totale (essendo pari al 78.7%).

Andiamo ora ad analizzare i residui in Figura 3.6.

L'istogramma ci fa notare una asimmetria dei residui verso sinistra ed anche il *normal probability plot* evidenzia un andamento ondulatorio.

Dal grafico dei residui verso i valori stimati possiamo dire che ci sono dei dati che portano a dubitare della distribuzione casuale degli errori.

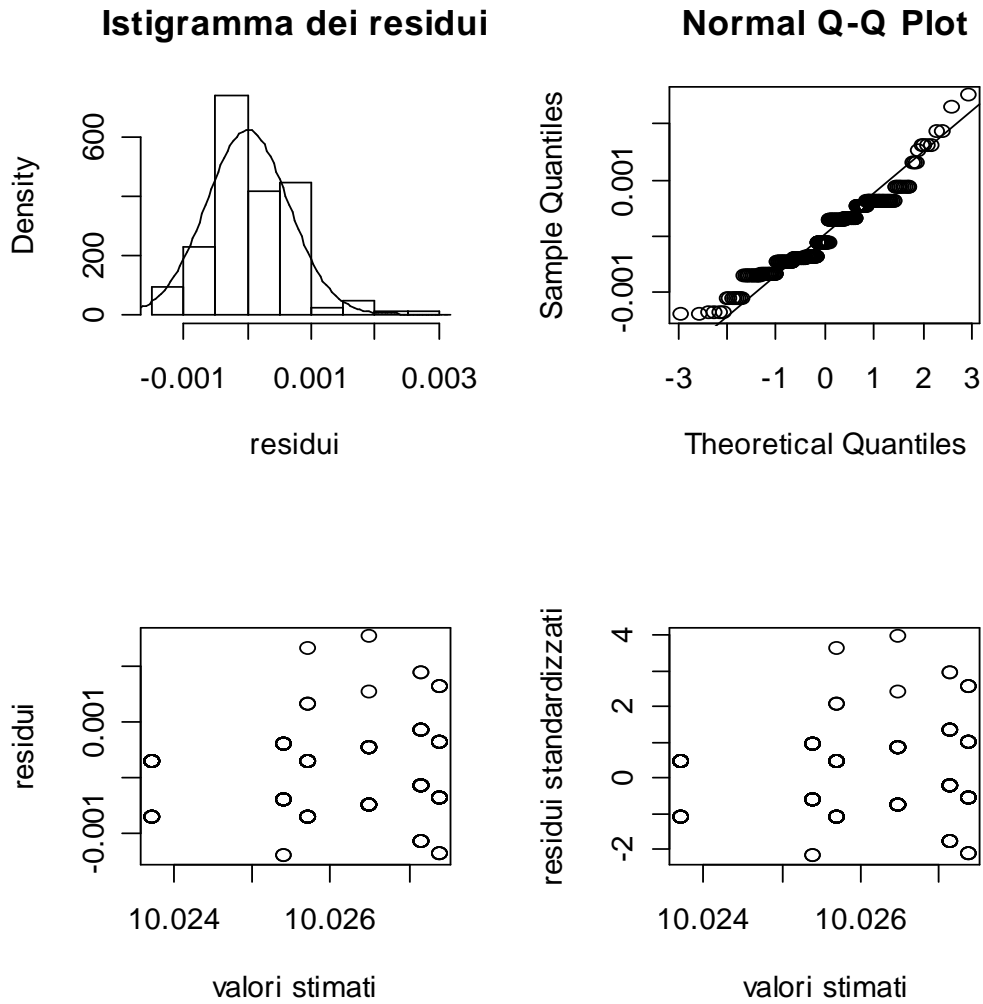


Figura 3.6: Diagnostiche grafiche sui residui del modello adattato al Codice 653

In definitiva nonostante l'adattamento ai dati del modello sembri essere adeguato, le diagnostiche grafiche evidenziano aspetti non del tutto soddisfacenti.

### 3.4 Codice 755

Il dataset del Codice 755 contiene 400 misurazioni suddivise questa volta in 8 campioni sempre di ampiezza 50 ciascuno. Le variabili rilevate sono:

- 'Foro' : variabile continua contenente le misure dei diametri;
- 'Gruppo' : variabile categoriale a 8 livelli indicante il gruppo di appartenenza;
- 'Trend' : variabile categoriale a 4 livelli che identifica l'andamento crescente o decrescente dei dati nel tempo. Tale variabile è stata introdotta guardando la distribuzione dei dati nella Figura 2.5;
- 'Intervento' : variabile categoriale a 4 livelli che indica se vi è stato o meno un intervento da parte dell'operatore sulla macchina. Eravamo partiti da 2 variabili distinte: una 'Intervento', dicotomica che indicava '0' assenza, '1' presenza di intervento sulla macchina; ed una seconda variabile 'Operatore' sempre categoriale che indicava l'operatore. Per semplificare l'analisi abbiamo accorpato le 2 variabili in maniera da avere '0' se non vi è intervento ed il numero identificativo dell'operatore se invece c'è un intervento.
- 'ENL' : variabile categoriale a 5 livelli che mostra il passaggio ad un nuovo processo quando si è avuto un intervento da parte dell'operatore sulla macchina. Abbiamo la presenza di 5 livelli in quanto in ogni singolo gruppo si è intervenuti al più 5 volte.



C'è inoltre da dire che non ci siamo soffermati per quanto riguarda la natura dell'intervento che l'operatore apporta alla macchina. Questo infatti può solo consistere in un aumento o diminuzione di materiale da asportare da parte della mola a seconda che il pezzo tenda ad essere stretto oppure largo. Ovviamente finché il pezzo è stretto si può rilavorare, quando inizia a divenire troppo largo ed esce dalla tolleranza andrà scartato.

Il valore target del processo non è fissato su un punto preciso, o meglio non abbiamo una tolleranza simmetrica. Il valore nominale è 7 mm, con le tolleranze abbiamo un range di 7.012 mm e 7.022 mm, quindi possiamo pensare che il livello nominale del processo sia collocato a metà di tale intervallo sul valore 7.017.

Come si è visto nel grafico in Figura 2.5 i dati hanno un andamento lineare a tratti ed a volte escono addirittura dai limiti di tolleranza. Per cogliere questo particolare movimento dei dati si è deciso di creare una variabile categoriale 'Trend' a 4 livelli. Questa, in particolare, assumerà valore '1' per le prime 59 osservazioni, valore '2' per i successivi 62 dati, valore '3' per i seguenti 202 ed infine abbiamo 64 misure alle quali viene attribuito il livello '4'.

Con riferimento sempre al grafico di Figura 2.5 possiamo ipotizzare che i dati non si distribuiscono secondo una legge Normale. Analizzando l'istogramma di Figura 3.7 siamo in grado di dire che le barre non sono molto simmetriche tra di loro, ma la curva di densità adattata è molto simile ad una distribuzione normale. Per quanto riguarda il diagramma a scatola con baffi possiamo dire che è abbastanza simmetrico. La mediana è leggermente spostata verso il 3° quartile, ma non molto da farci dubitare della normalità. Infine, abbiamo la presenza di 3 osservazioni anomale.

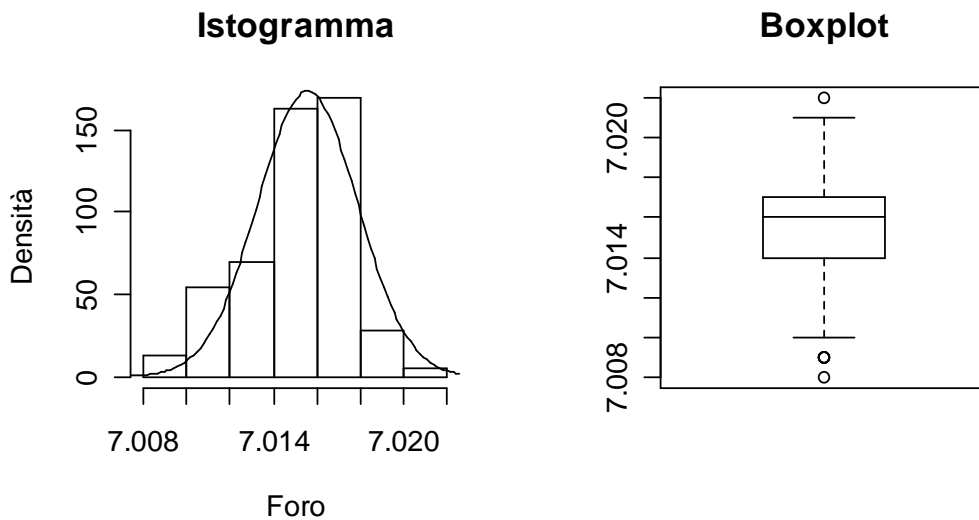


Figura 3.7: Istogramma con densità e boxplot dei dati del Codice 755

Il *normal probability plot* (Figura 3.8) mostra un andamento fluttuante attorno alla retta con la presenza di 3 curvature e di code pesanti (esterne all'intervallo  $[-2 : 2]$ ). In particolare si nota una dispersione soprattutto nella coda inferiore.

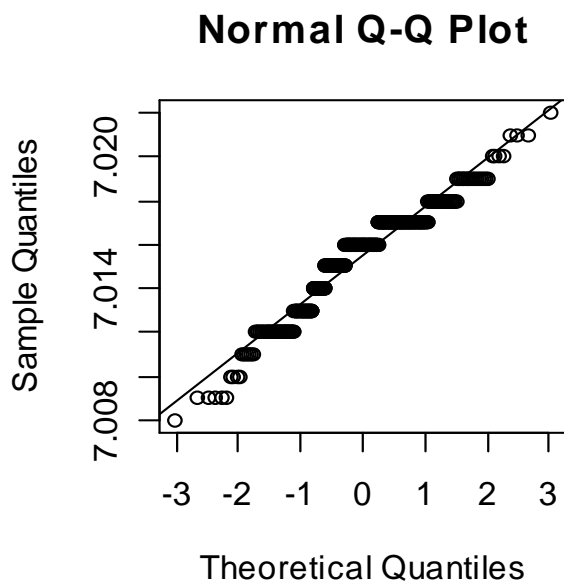


Figura 3.8: Normal probability plot dei dati del Codice 755

Applicando il *test di Shapiro-Wilk* otteniamo il seguente valore con il relativo livello di significatività osservato:

$$W = 0.9454 \qquad \alpha^{oss} = 5.58e - 11$$

Questi valori portano dunque a rifiutare l'ipotesi che i dati si distribuiscano come una variabile casuale Normale.

La verifica dell'ipotesi di normalità è stata eseguita anche per i singoli gruppi, ma tale ipotesi viene sempre rifiutata.

Analizziamo ora le distribuzioni condizionate per mezzo dei diagrammi a scatola con baffi (Figura 3.9).

Se la variabile 'Intervento' assume valore '0', cioè non c'è stato nessun intervento da parte dell'operatore sulla macchina, il boxplot risulta simmetrico e con una variabilità più elevata. I suoi baffi infatti sono estesi e con un'osservazione anomala per lato.

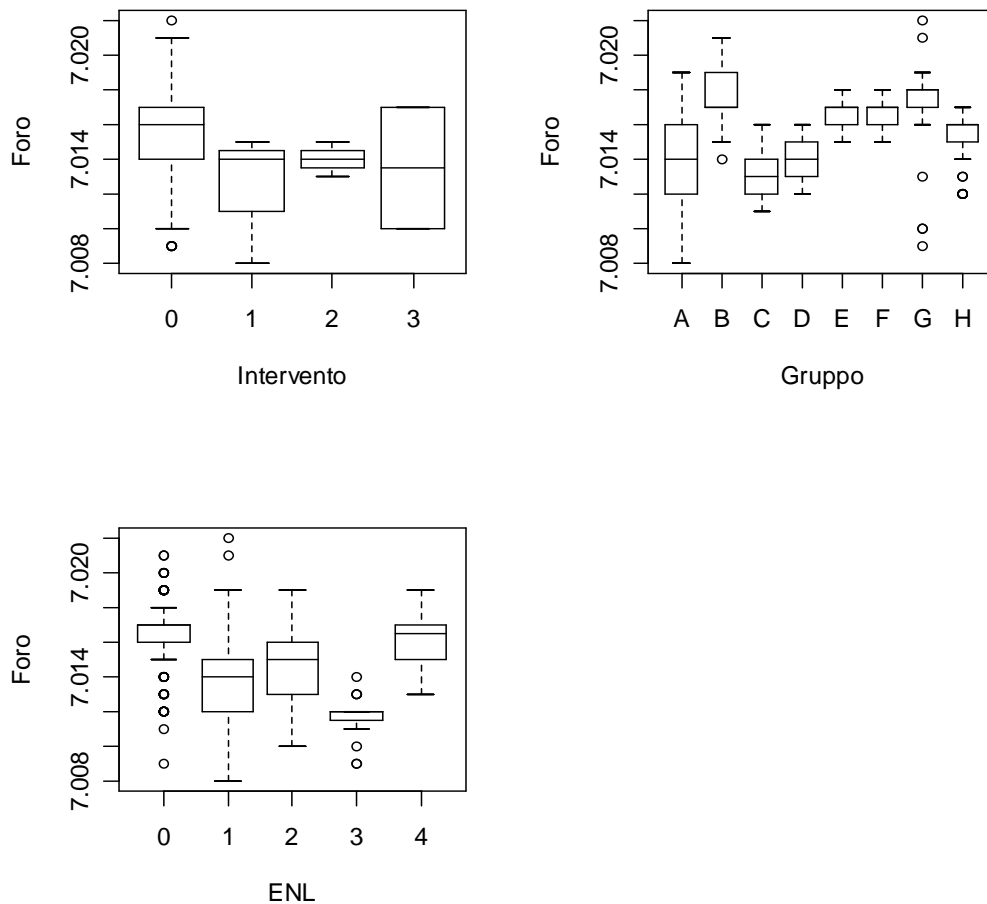


Figura 3.9: Boxplot della variabile 'Foro' rispetto alle altre variabili

Per 'Intervento = 1' si riscontra un'asimmetria verso il primo quartile, mentre quando la variabile assume il valore '2' il grafico è simmetrico e con pochissima variabilità, infine per il valore '3' otteniamo una scatola simmetrica e senza baffi. Va fatta una considerazione sulla numerosità dei gruppi della variabile in quanto la quasi totalità dei dati, per la precisione 387, si riferiscono al primo livello, mentre negli altri abbiamo rispettivamente 8, 3 e 2 osservazioni. Quindi nel comparare i boxplot bisogna usare mostrare molta cautela.

Nella variabile 'Gruppo', fatta eccezione per i gruppi E e F, abbiamo una completa diversità di grafici. I primi due sono molto più variabili rispetto ai

restanti e nel gruppo G abbiamo il maggior numero di osservazioni anomale.

Infine i livelli individuati nella variabile 'ENL' producono scatole con variabilità, simmetria e posizioni parecchio diverse tra di loro. Anche in questo caso le osservazioni non sono ripartite in maniera eguale nei gruppi. Tuttavia per la modalità 'ENL = 0', che presenta il numero di osservazioni maggiori, la scatola è molto ristretta ma mostra parecchi valori anomali (sono questi infatti che ci portano al rifiuto dell'ipotesi di normalità).

In Figura 3.10 e 3.11 sono riportati rispettivamente il grafico totale e quello parziale dell'autocorrelazione dei dati:

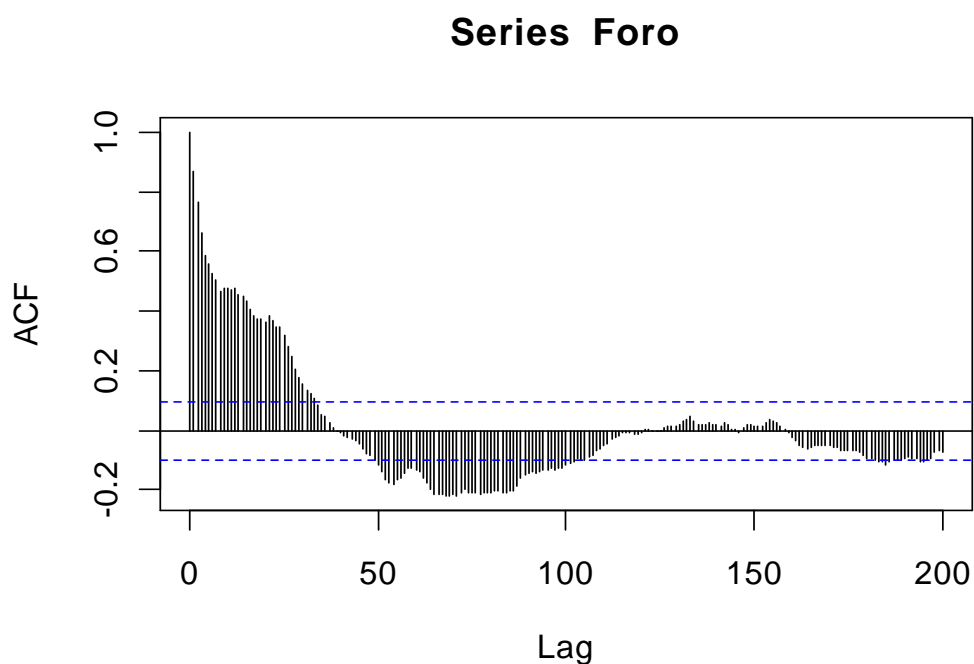


Figura 3.10: Grafico della autocorrelazione totale, Codice 755

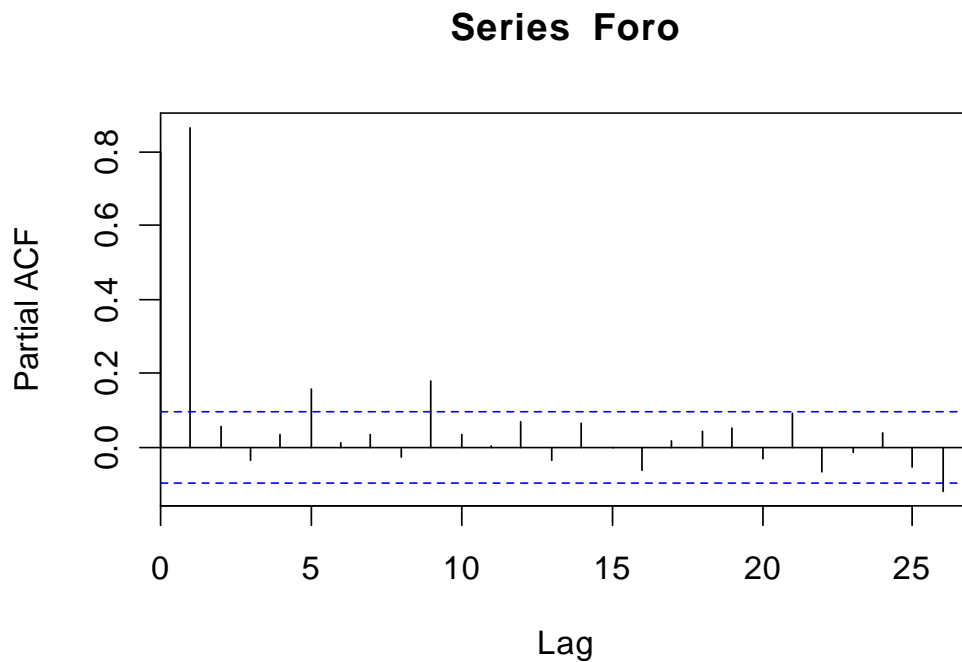


Figura 3.11: Grafico della autocorrelazione parziale, Codice 755

Si può notare come le osservazioni siano tra loro fortemente correlate. Abbiamo la presenza di “onde” che non hanno però ampiezza e lunghezza costante, positive e vicine ad 1 all’inizio vanno poi decrescendo ed invertono il loro segno. Questo andamento è tipico delle serie non stazionarie. Il correlogramma ci dice che la lunghezza media di queste onde è di 50 periodi circa, ciò corrisponde alla numerosità nei gruppi. Di seguito in Figura 3.12, riportiamo i grafici di autodispersione fino al terzo periodo della variabile ‘Foro’.

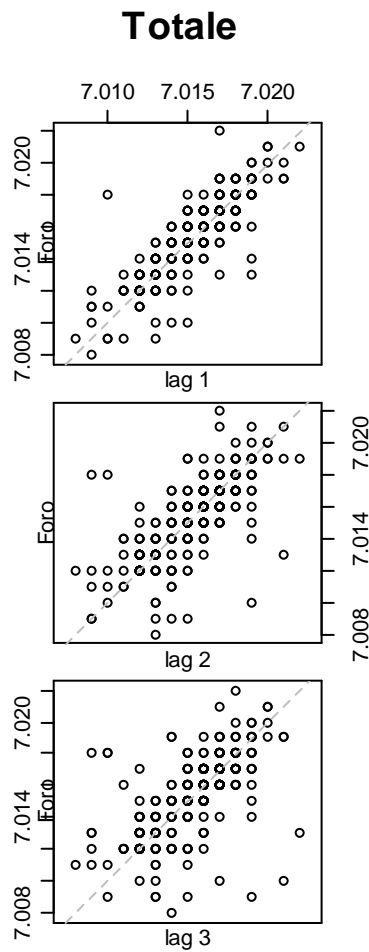


Figura 3.12: Grafico di autodispersione, Codice 755

Si vede un forte legame lineare che associa le variabili a ritardo 1; al secondo ritardo il legame è ancora presente ma i dati si scostano un po' dalla retta. Infine a ritardo 3 cominciamo ad avere dati più dispersi.

Disponiamo quindi di 5 variabili, di cui una continua e 4 di tipo fattore, e 400 osservazioni. Queste sono fortemente autocorrelate e non seguono un andamento normale.

Per quanto riguarda il dataset del Codice 755 è stato costruito un modello avente come risposta la variabile 'Foro' e come esplicative tutte le altre.

Otteniamo così un modello con 17 parametri le cui analisi sono riportate di seguito:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	7.0136214	0.0003396	20652.242	< 2e-16	***
Trend2	-0.0016829	0.0005403	-3.115	0.001981	**
Trend3	-0.0006402	0.0009315	-0.687	0.492293	
Trend4	0.0011532	0.0010768	1.071	0.284907	
Intervento1	-0.0010603	0.0005434	-1.951	0.051771	.
Intervento5	0.0010879	0.0008777	1.239	0.215941	
Intervento22	-0.0037739	0.0010708	-3.524	0.000476	***
GruppoB	0.0053786	0.0005956	9.031	< 2e-16	***
GruppoC	0.0020326	0.0006717	3.026	0.002647	**
GruppoD	0.0015630	0.0010570	1.479	0.140034	
GruppoE	0.0035388	0.0010822	3.270	0.001173	**
GruppoF	0.0038388	0.0010822	3.547	0.000438	***
GruppoG	0.0040473	0.0010795	3.749	0.000205	***
GruppoH	0.0007342	0.0012091	0.607	0.544096	
ENL1	-0.0007502	0.0002702	-2.776	0.005774	**
ENL2	-0.0005523	0.0002948	-1.873	0.061778	.
ENL3	-0.0029171	0.0006818	-4.279	2.38e-05	***
ENL4	0.0027153	0.0004846	5.603	4.05e-08	***

Residual standard error: 0.001468 on 382 degrees of freedom  
 Multiple R-Squared: 0.6107, Adjusted R-squared: 0.5934  
 F-statistic: 35.26 on 17 and 382 DF, p-value: < 2.2e-16

In questo caso non tutte le categorie delle variabili fattoriali assumono un effetto significativo.

L'  $R^2$  associato al modello è pari a 0.6107, quindi con questo modello spieghiamo il 61.07% dalla varianza totale.

L'analisi dell'istogramma dei residui in Figura 3.13 ci suggerisce che essi assumano un andamento approssimativamente normale.



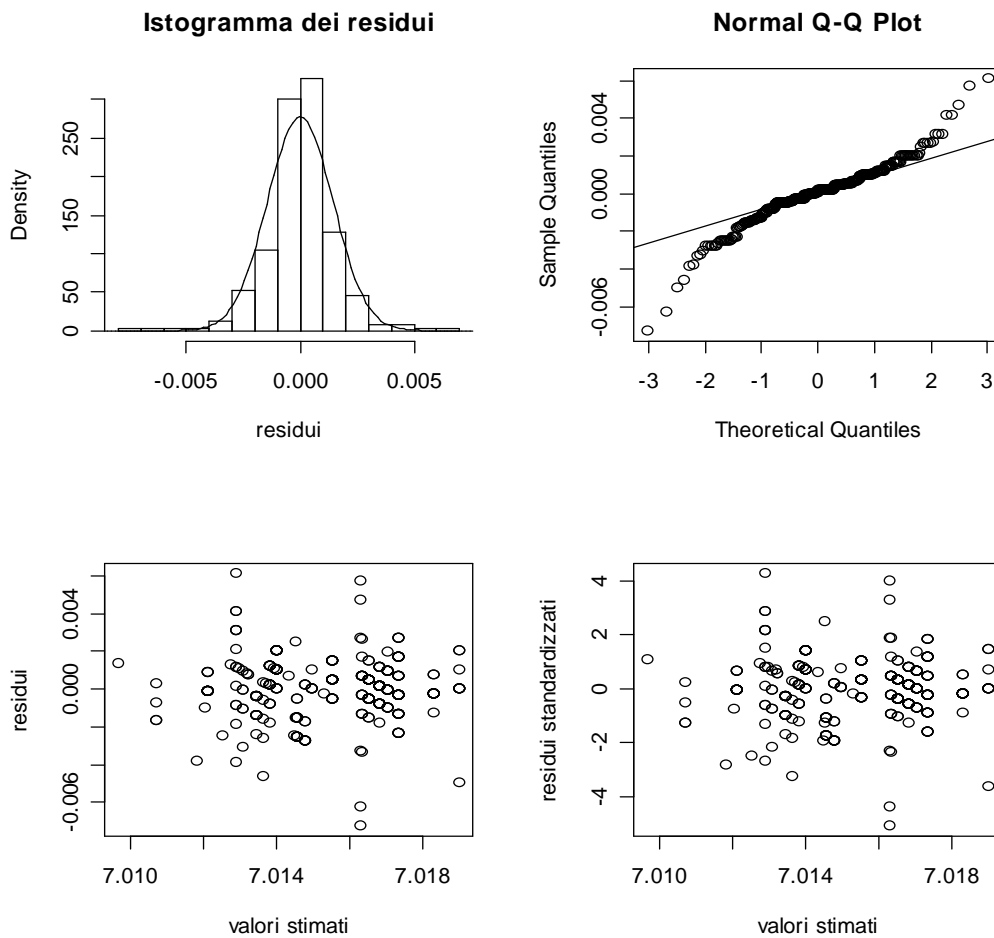


Figura 3.13: Diagnostiche grafiche sui residui del modello adattato al Codice 755

Il *normal probability plot* presenta un andamento abbastanza soddisfacente nella parte centrale, ma verso le code della distribuzione, fuori dell'intervallo  $[-2;2]$ , si evidenziano delle curvature. Questo può essere dovuto al fatto che i residui sono molto più grandi in valore assoluto dei corrispondenti valori attesi sotto l'ipotesi di normalità.



## CAPITOLO 4

### *Le carte di controllo*

L'SPC è lo strumento principale per raggiungere la stabilità e per migliorare un processo produttivo attraverso la riduzione della variabilità dei fattori coinvolti nella produzione.

Le carte di controllo rappresentano invece la soluzione più semplice per realizzare delle procedure di controllo statistico di processo e i loro obiettivi principali sono:

- ridurre la variabilità del processo
- monitorare e controllare il processo
- stimare i parametri del prodotto e del processo

La realizzazione di una carta di controllo prevede la costruzione di un grafico riportante i valori della statistica di controllo di interesse. Nel grafico vengono disegnate una *Linea Centrale (Central Line, LC)* la quale identifica il valor medio della qualità misurata, ed altre due linee a distanza simmetrica da questa dette *Limite di Controllo Superiore (Upper Control Limit, UCL)* e *Limite di Controllo Inferiore (Lower Control Limit, LCL)*.

Le costanti UCL ed LCL identificano quindi i valori critici con cui la statistica viene confrontata e, se il processo in esame è in controllo statistico, tale statistica deve cadere all'interno di detti limiti con un livello di probabilità prefissato.

Le carte di controllo possono perciò essere assimilate ad una successione sequenziale di test di verifica di ipotesi del tipo:

$$\begin{cases} H_0: & \text{processo in controllo} \\ H_1: & \text{processo fuori controllo} \end{cases}$$

dove per ogni istante di tempo si procede a verificare se è vera l'ipotesi nulla. Se la statistica cade nella regione di rifiuto si conclude che il

processo è *fuori controllo* lanciando quindi un allarme. Abbiamo invece un falso allarme quando la statistica segnala un allarme e tuttavia il processo è in controllo. Si parlerà di *efficienza* di una carta di controllo quando minimizzerà il ritardo nella segnalazione di un effettivo cambiamento sotto un vincolo sul numero medio di falsi allarmi.

Vi sono diversi tipi di carte di controllo a seconda della caratteristica di processo che si vuole sorvegliare. Una prima differenza è data dal tipo di variabile di interesse: se essa è una variabile quantitativa e quindi misurabile fisicamente, costruiremo carte di controllo per variabili, se invece la variabile è di tipo qualitativo si costruiranno carte di controllo per attributi. Nelle nostre analisi utilizzeremo carte di controllo per variabili in quanto poniamo sotto controllo i diametri dei fori di due ingranaggi.

Un'altra differenza nelle carte di controllo per variabili è data dalle carte con o senza memoria a seconda che la statistica in esame tenga conto o meno delle precedenti osservazioni.

#### 4.1. Le carte Shewhart

Le carte di controllo di tipo Shewhart (Shewhart, 1931) sono delle carte senza memoria, cioè durante l'analisi si tiene conto solamente dell'osservazione corrente.

Si individua una certa caratteristica del prodotto  $X$  e la si misura in diversi istanti temporali. Si calcola la statistica di controllo  $w$  che misura tale caratteristica:

$$w = g(\underline{x}) \quad 4.1$$

La statistica di controllo è tale da avere media e deviazione standard:

$$\begin{aligned} E[w] &= \mu_w \\ \sqrt{V[w]} &= \sigma_w \end{aligned} \quad 4.2$$

La regola di decisione si basa sull'ultimo campione estratto per mezzo della funzione:

$$w_t = g(\underline{x}_t), \quad 4.3$$

i limiti di controllo sono fissi per ogni istante temporale e si saggia il seguente sistema di ipotesi:

$$\begin{cases} H_0 : \mu = \mu_w \\ H_1 : \mu \neq \mu_w \end{cases}.$$

La linea centrale LC è posta pari a  $\mu_w$  che può rappresentare il valore nominale della caratteristica, può essere basata sull'esperienza passata oppure stimata a partire dai dati. Da essa si calcolano i limiti di controllo:

$$\begin{aligned} LCL &= \mu_w - L\sigma_w \\ UCL &= \mu_w + L\sigma_w \end{aligned} \quad 4.4$$

dove  $L$  rappresenta la distanza dei limiti dalla linea centrale. Per il calcolo della costante  $L$  possiamo usare diversi approcci.

Un primo criterio dice che dato un  $\alpha$ , l'errore di I° tipo che consiste nel rifiutare  $H_0$  quando in realtà è vera ( Rischio del Produttore),  $L$  si sceglie in maniera che sotto l'ipotesi nulla:

$$P[w \notin (LCL, UCL)] = \alpha \quad 4.5$$

Se in particolare la distribuzione di  $w$  è simmetrica  $L$  si sceglie in modo che:

$$P[w \leq \mu_w - L\sigma_w | H_0] + P[w \geq \mu_w + L\sigma_w | H_0] = \alpha \quad 4.6$$

Un secondo approccio prevede di considerare l'ARL (Average Run Length), ovvero il tempo medio di attesa per segnalare un allarme. Se il processo è in controllo abbiamo che il valore di  $\alpha$  è quello visto nella 4.5 sotto  $H_0$  e l'ARL<sub>0</sub>, cioè l'ARL in controllo, risulta:

$$ARL_0 = E[N] = \frac{1}{\alpha} \quad 4.7$$

dove  $N$  è una variabile Geometrica di parametro  $\alpha$ .

Se il processo è invece fuori controllo  $\beta$ , l'errore di II° tipo che si commette rifiutando  $H_0$  quando è falsa (Rischio del Cliente), risulta pari a:

$$\beta = P[w \in (LCL, UCL)] \quad 4.8$$

Sotto l'ipotesi alternativa, otteniamo quindi un  $ARL_1$  fuori controllo pari a:

$$ARL_1 = E[N] = \frac{1}{1-\beta}. \quad 4.9$$

Si fissa quindi l'ARL in controllo, si determina  $\alpha$  e di conseguenza  $L$ . Se  $\alpha$  diminuisce i limiti di controllo diventano più ampi e  $\beta$  aumenta, viceversa se  $\alpha$  aumenta i limiti di controllo diventano più stretti e  $\beta$  diminuisce.

Un'altra proposta per la scelta di  $L$  si basa sulla regola del '3 - Sigma'. Tale regola fissa

$$L = \pm 3$$

in quanto basandoci sul teorema del limite centrale se abbiamo che

$$w \sim N(\mu_w, \sigma_w^2)$$

il  $100(1-\alpha)\%$  dei valori di  $w \in \left( \mu_w \pm Z_{\frac{\alpha}{2}} \sigma_w \right)$ . In particolare

l'intervallo  $\mu_w \pm 3\sigma_w$  contiene il 99.7% dei valori di  $w$ . Il limite del '3-Sigma' può essere mediamente superato 3 volte su mille. Accade cioè raramente che un campione porti ad un valore di  $w$  esterno a tali valori quando la media di  $w$  è  $\mu_w$ .

Nello svolgimento della tesi il valore di  $L$  verrà scelto secondo quest'ultimo approccio, quindi per tutte le carte di controllo che faremo  $L$  sarà posto a  $\pm 3$ .

#### 4.1.1 Misure singole

Dato che siamo in presenza di osservazioni singole se analizziamo la media del processo i parametri 4.1 e 4.2 diverranno:

$$\begin{aligned} w &= x_i \\ E[w] &= E[x_i] = \mu_0 \\ \sigma_w &= \sqrt{V[w]} = \sqrt{V[x_i]} = \sigma \end{aligned} \quad 4.10$$

la media sarà quindi stimata attraverso la media campionaria dei dati mentre la stima della deviazione standard verrà basata sulle escursioni mobili campionarie

$$MR(2) = |x_j - x_{j-1}| \quad 4.11$$

Da ciò risulta:

$$\begin{aligned} \mu_w &= E[MR(2)] = d_2 \sigma \\ \sigma_w &= \sqrt{V[MR(2)]} = \sqrt{d_3^2 \sigma^2} = d_3 \sigma \end{aligned} \quad 4.12$$

dove  $d_2$  e  $d_3$  sono delle costanti tabulate in funzione di  $n$ . Uno stimatore corretto per  $\sigma_x$  è dato da:

$$\hat{\sigma}_x = \frac{\hat{E}[MR(2)]}{d_2} = \frac{\overline{MR(2)}}{d_2} \quad 4.13$$

dove

$$\overline{MR(2)} = \frac{\sum_{i=1}^n |x_i - x_{i-1}|}{n - 1}. \quad 4.14$$

Ponendo

$$\mu_w \pm 3\sigma_w \Rightarrow d_2\sigma \pm 3d_3\sigma$$

i limiti di controllo della 4.4 risultano:

$$\begin{aligned}
LCL &= \overline{MR}(2) - 3\frac{d_3}{d_2}\overline{MR}(2) \\
LC &= \overline{MR}(2) \\
UCL &= \overline{MR}(2) + 3\frac{d_3}{d_2}\overline{MR}(2)
\end{aligned}
\tag{4.15}$$

dove per  $n = 2$  la costante  $d_2$  assume il valore 1.128 e  $d_3$  vale 0.28.

#### 4.2. Le carte EWMA

Le carte di controllo di *Exponentially Weighted Moving Average*, EWMA, (Roberts, 1959), fanno parte delle carte di controllo che usano l'informazione proveniente da campioni precedenti in aggiunta a quella proveniente dal campione attuale. La memoria delle carte EWMA è definita come infinita e non uniforme nel tempo.

Le carte EWMA possiedono notevoli proprietà:

- adatte per l'individuazione di salti di entità limitata;
- robuste alla non normalità delle singole misure;
- considerano la correlazione tra osservazioni passate;
- evidenziano trend (spostamento lento ma sistematico verso l'alto o il basso della statistica di controllo) di deriva anche con poche osservazioni;
- ideali per le carte sulle misure singole;
- necessitano di un solo grafico.

La statistica di controllo EWMA consiste in una media mobile ponderata delle osservazioni passate:

$$w_t = \sum_{i=0}^{t-1} \lambda(1-\lambda)^i w_{t-i} + (1-\lambda)^t w_0 \tag{4.16}$$

dove  $w_0$  è il valore iniziale e  $0 < \lambda \leq 1$  è una costante che rappresenta il peso assegnato a ciascuna osservazione. Tali pesi decrescono in progressione geometrica e la loro somma è pari all'unità.



I valori rappresentano direttamente una stima del parametro sottoposto a controllo e variano attorno alla linea centrale con piccole fluttuazioni quando il processo è in controllo.

#### 4.2.1. Misure singole

Se si dispone di misure singole il valor medio e la varianza della statistica di controllo quando il processo è in controllo risultano:

$$\begin{aligned}\mu_w &= E[w_t] = \mu_0 \\ \sigma_w^2 &= \sigma_0^2 \left( \frac{\lambda}{2-\lambda} \right) [1 - (1-\lambda)^{2t}]\end{aligned}\quad 4.17$$

dal momento che  $1 - (1-\lambda)^{2t} \rightarrow 1$  abbiamo che per  $t \rightarrow \infty$  la varianza diviene:

$$\sigma_w^2 = \sigma_0^2 \left( \frac{\lambda}{2-\lambda} \right). \quad 4.18$$

Saremo quindi in presenza di due tipi di limiti:

- limiti variabili per valori piccoli di t:

$$\begin{aligned}LC &= \mu_0 \\ (LCL, UCL) &= \mu_0 \pm L\sigma_0 \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]}\end{aligned}\quad 4.19$$

- limiti asintotici al crescere di t:

$$(LCL, UCL) = \mu_0 \pm L\sigma_0 \sqrt{\frac{\lambda}{2-\lambda}} \quad 4.20$$

Se la media e la deviazione standard del processo sono incognite i limiti asintotici divengono:

$$\begin{aligned}LC &= \bar{x} \\ (LCL, UCL) &= \bar{x} \pm L\hat{\sigma} \sqrt{\frac{\lambda}{2-\lambda}}\end{aligned}\quad 4.21$$

dove:

$$\bar{x} = \frac{\sum x_i}{n}, \quad \hat{\sigma} = \frac{\overline{MR}(2)}{d_2} \quad 4.22$$

#### 4.2.2. Campioni di ampiezza $n > 1$

Se si dispone di campioni di ampiezza  $n > 1$  dobbiamo dividere la varianza per  $\sqrt{n}$  ed i limiti 4.19 e 4.20 cambiano in:

- limiti variabili:

$$(LCL, UCL) = \mu_0 \pm L \frac{\sigma_0}{\sqrt{n}} \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]} \quad 4.23$$

- limiti asintotici:

$$(LCL, UCL) = \mu_0 \pm L \frac{\sigma_0}{\sqrt{n}} \sqrt{\frac{\lambda}{2-\lambda}} \quad 4.24$$

Se la media e la deviazione standard sono ignote i limiti asintotici diventano:

$$(LCL, UCL) = \bar{x} \pm L \frac{\hat{\sigma}}{\sqrt{n}} \sqrt{\frac{\lambda}{2-\lambda}} \quad 4.25$$

dove:

$$\bar{x} = \frac{\sum \bar{x}_i}{n}, \quad \hat{\sigma} = \frac{\bar{R}}{d_2} \quad \text{oppure} \quad \hat{\sigma} = \frac{\bar{s}}{c_4} \quad 4.26$$

$\bar{R}$  ed  $\bar{s}$  sono le medie rispettivamente dei range e delle deviazioni standard campionarie, mentre  $d_2$  e  $c_4$  sono costanti che dipendono dalla numerosità campionaria.

I valori di  $\lambda$  e di  $L$  vengono scelti in relazione all'  $ARL$  e studi su questo approccio sono stati fatti da Crowder (1987) e Lucas e Saccucci (1990).

Per progettare una procedura ottimale si specificano l'  $ARL_0$  e la minima ampiezza del salto di media del processo che si vuole segnalare, in questa maniera si vanno a selezionare valori di  $\lambda$  e di  $L$  che garantiscono il minimo valore dell'  $ARL$  fuori controllo.

Scegliendo un valore di  $\lambda \in (0.05, 0.25)$  otteniamo uno schema efficiente nei confronti di cambiamenti di piccola entità del processo, mentre per valori di  $\lambda$  grandi abbiamo uno schema più efficiente nei confronti di cambiamenti del processo di grande entità. La costante  $\lambda$  risulta quindi un parametro di liscio che aumentando dà più peso all'ultima osservazione, laddove per valori piccoli si dà maggior enfasi al passato.

#### 4.3 Carte di controllo per dati autocorrelati

Le assunzioni di base per l'uso delle carte di controllo prevedono che i dati siano indipendenti e normalmente distribuiti. Tuttavia in molte situazioni reali l'ipotesi di incorrelazione non è soddisfatta e la performance delle carte Shewhart applicata ai dati risulta insoddisfacente.

Una tecnica utile consiste nell'individuare dei modelli (tipicamente di serie storica) da adattare ai dati per eliminare la presenza di autocorrelazione ed applicare una carta Shewhart ai residui di tali modelli. I residui prendono quindi la forma:

$$e_t = y_t - \hat{y}_t$$

dove  $\hat{y}_t$  sono i valori adattati secondo un conveniente modello di serie storica. Per un buon adattamento l'andamento dei punti deve essere casuale all'interno dei limiti.

A differenza della Shewhart la carta EWMA può essere direttamente applicata sui dati autocorrelati. In questo caso si procede ad applicare una carta Shewhart sugli errori di previsione ad un passo dati da:

$$e_t = y_t - w_{t-1}$$

dove il valore  $w_{t-1}$  indica il valore della statistica EWMA al tempo  $t-1$ , ovvero la previsione di  $\mu_t$ .

Minimizzando la somma dei quadrati di tali residui riusciamo ad ottenere una stima ottimale per il parametro  $\lambda$  (ciò assicura che i residui siano incorrelati). Tale stima, che chiameremo  $\lambda^{opt}$ , identifica il valore da utilizzare per il disegno della carta.

Quando l'EWMA costituisce un buon previsore ad un passo è possibile utilizzare  $w_t = \hat{x}_{t+1}$  come linea centrale per la costruzione di una carta di controllo per il periodo  $t+1$  ed i limiti varieranno nel tempo secondo la seguente formula:

$$\begin{aligned} UCL_{t+1} &= w_t + Z_{\frac{\alpha}{2}} \sigma_p \\ LCL_{t+1} &= w_t - Z_{\frac{\alpha}{2}} \sigma_p \end{aligned}$$

dove  $\hat{\sigma}_p^2$ , nell'ipotesi che il parametro  $\lambda$  venga scelto a partire da  $n$  osservazioni, si ottiene dividendo la somma degli errori di previsione al quadrato, ottenuti con il  $\lambda^{opt}$ , per  $n$ .

Questa procedura prende il nome di carta di controllo EWMA a linea centrale mobile ed è molto utile in quanto offre informazioni sulla dinamica del processo.

#### 4.4 Applicazione delle carte di controllo al Codice 653

L'analisi esplorativa ha evidenziato che i campioni provengono plausibilmente da processi diversi.

Per sorvegliare la media del processo costruiamo una carta  $\bar{X}$  per ogni gruppo. Dalle carte di controllo costruite si nota che gli andamenti nei diversi gruppi sono piuttosto simili; quindi nella Figura 4.1 riportiamo solamente il grafico del gruppo A per un'analisi descrittiva della carta di controllo (i grafici relativi agli altri gruppi sono riportati nell'Appendice A).

I limiti di controllo per il gruppo A, calcolati con le formule viste nel capitolo precedente, sono quelli riportati in rosso e risultano:

$$LCL = 10.02418$$

$$LC = 10.0257$$

$$UCL = 10.02722$$

Il grafico è il seguente (con le linee verdi sono evidenziati i limiti di tolleranza):

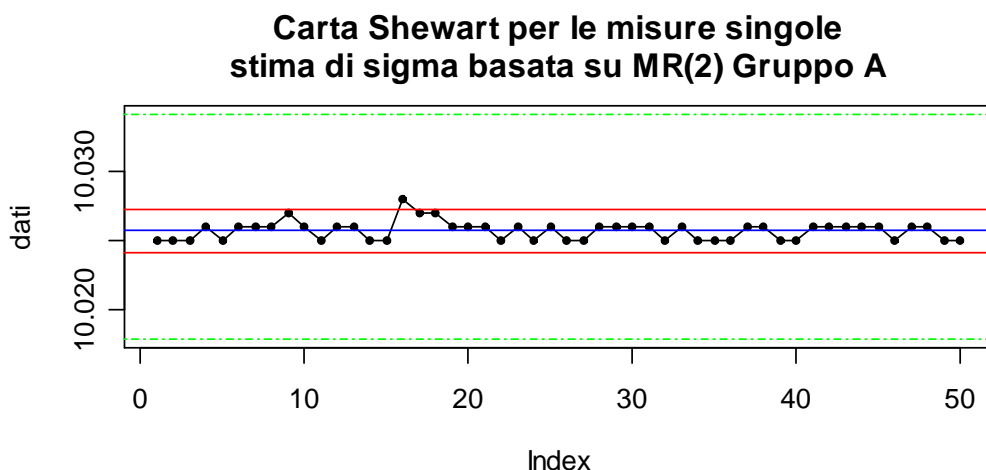


Figura 4.1: Carta  $\bar{x}$  per il Gruppo A, Codice 653

Si può vedere come un solo valore cada fuori dai limiti di controllo e anche per gli altri gruppi tutte le osservazioni stanno all'interno dei limiti di controllo. Per avere un'idea anche della variabilità entro i gruppi

suddividiamo i dati in sottogruppi “mobili” di numerosità 2 (ogni elemento viene confrontato con il successivo in ordine temporale) ed applichiamo una carta sulle escursioni mobili (Figura 4.2):

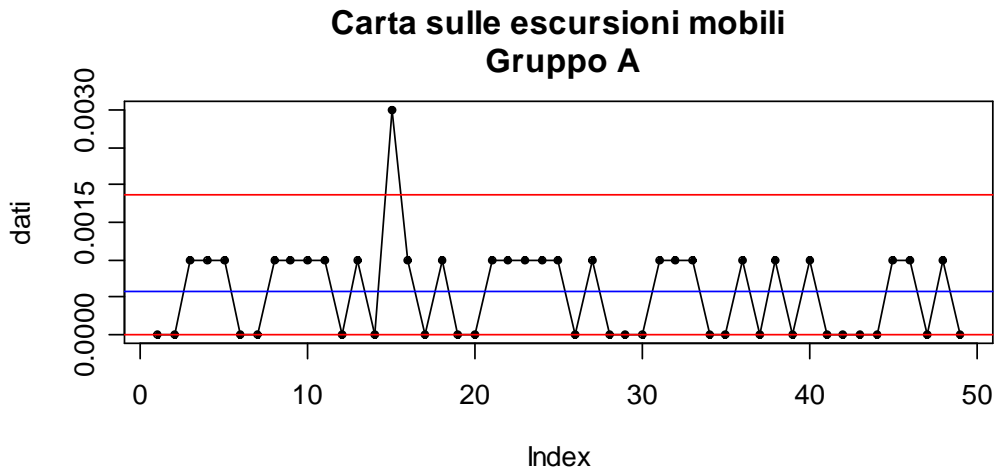


Figura 4.2: Carta sulle escursioni mobili per il Gruppo A, Codice 653

dalla quale si nota un salto tra le osservazioni 15 e 16 e un sostanziale andamento ciclico.

Data la presenza di autocorrelazione (nell’appendice A sono riportati i grafici delle osservazioni divisi per ogni gruppo con il rispettivo  $\lambda^{ott}$ ) disegniamo una carta di controllo di tipo Shewhart sugli errori di previsione ad un passo (Figura 4.3) avente come linea centrale 0 e limiti di controllo  $\pm 3\sigma_p$ , che nel caso in esame sono:

$$LCL = -0.00204347$$

$$UCL = 0.00204347$$

### Carta sugli errori di previsione ad un passo Totale

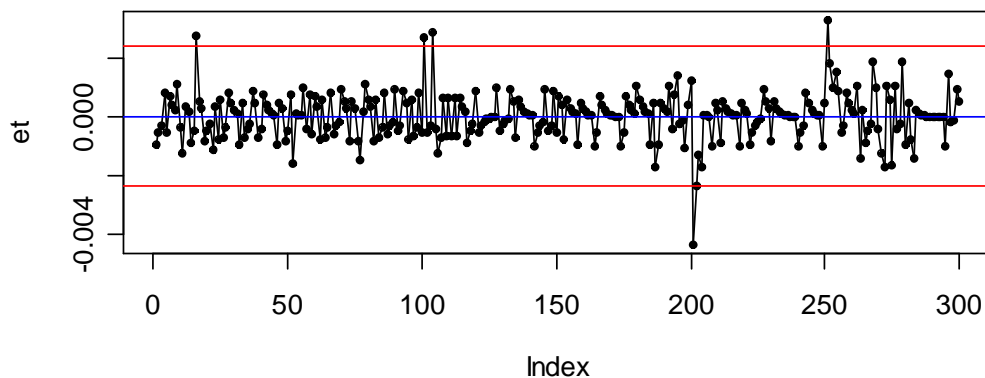


Figura 4.3: Carta sugli errori di previsione ad un passo, Codice 653

Il grafico mostra la presenza di diversi fuori controllo.

Su di un altro grafico (Figura 4.4) infine riportiamo le previsioni EWMA ad un passo mediante un carta avente con linea centrale il valore della statistica EWMA.

### Carta per le previsioni ad un passo EWMA center line - Totale

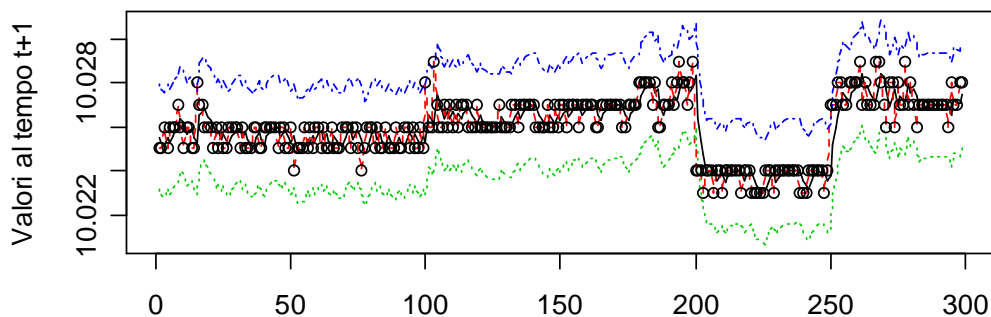


Figura 4.4: Carta per le previsioni ad un passo, Codice 653

Dalla Figura 4.4 si nota che viene riproposta la differenza tra i campioni, ciò è dovuto al fatto che il  $\lambda^{ott}$  è piuttosto grande e pari a 0.45.

L'adattamento del modello di regressione lineare ha mostrato che il diametro del foro non risulta influenzato solo dalla variabilità naturale ed intrinseca del processo ma anche da altri fattori di variabilità.

Sono state pertanto applicate delle carte di controllo EWMA ai residui dei modelli adattati nel capitolo precedente (Figura 4.5):

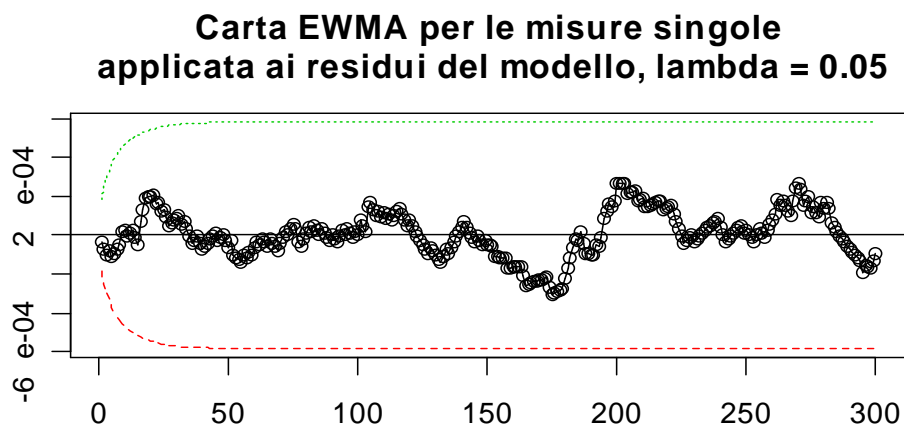


Figura 4.5: Carta EWMA applicata ai residui del modello adattato al Codice 653

Otteniamo uno strumento per il controllo della variabile aggiustata per l'effetto 'Gruppo' ed in questo caso il  $\lambda^{ott}$  assume valore 0.05, daremo così più peso alle osservazioni lontane.

La Figura mostra che nell'ipotesi in cui sia possibile sorvegliare l'effetto della variabile 'Gruppo' sulla variabilità totale del diametro questo risulta stabile in media nel tempo.



#### 4.5 Applicazione delle carte di controllo al Codice 755

Vista la forte differenza in media dei gruppi per il Codice 755 le carte di controllo  $\bar{X}$  per la media ed  $R$  sulla variabilità esibiscono parecchi valori fuori controllo. Anche per il Codice 755, data la correlazione seriale, si procede all'applicazione della Shewhart agli errori di previsione a un passo e alla EWMA *center line*.

I risultati sono apprezzabili, per il grafico sugli errori di previsione (Figura 4.6) infatti abbiamo che le osservazioni sono abbastanza concentrate attorno alla linea centrale e rileviamo la presenza di pochi fuori controllo dai limiti dati da:

$$LCL = -0.00354871$$

$$LC = 0$$

$$UCL = 0.00354871$$

#### Carta sugli errori di previsione ad un passo

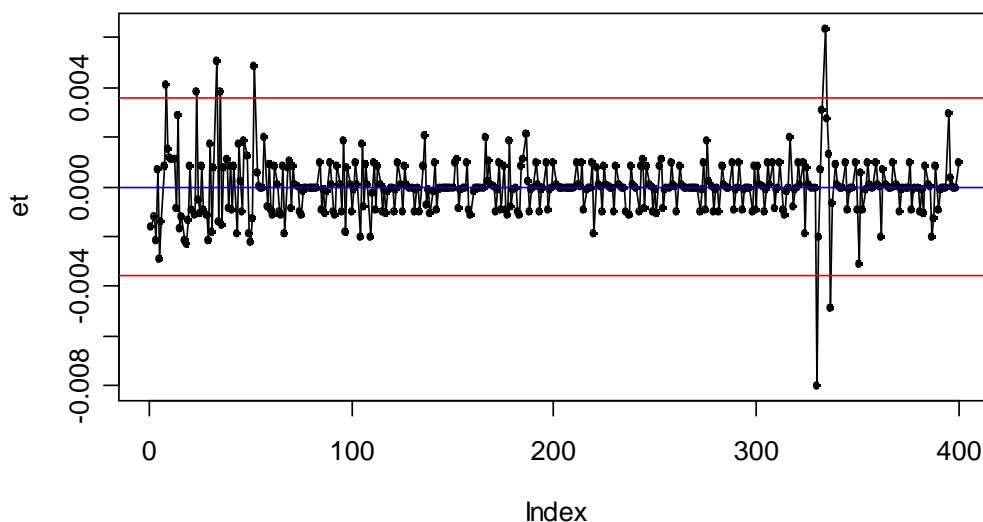


Figura 4.6: Carta sugli errori di previsione ad un passo, Codice 755

Per quanto riguarda le previsioni ad un passo (Figura 4.7) il grafico presenta un andamento oscillante come visto precedentemente. Il  $\lambda^{ott}$  in

questo caso vale 0.875 e quindi diamo più peso alle osservazioni vicine nel tempo.

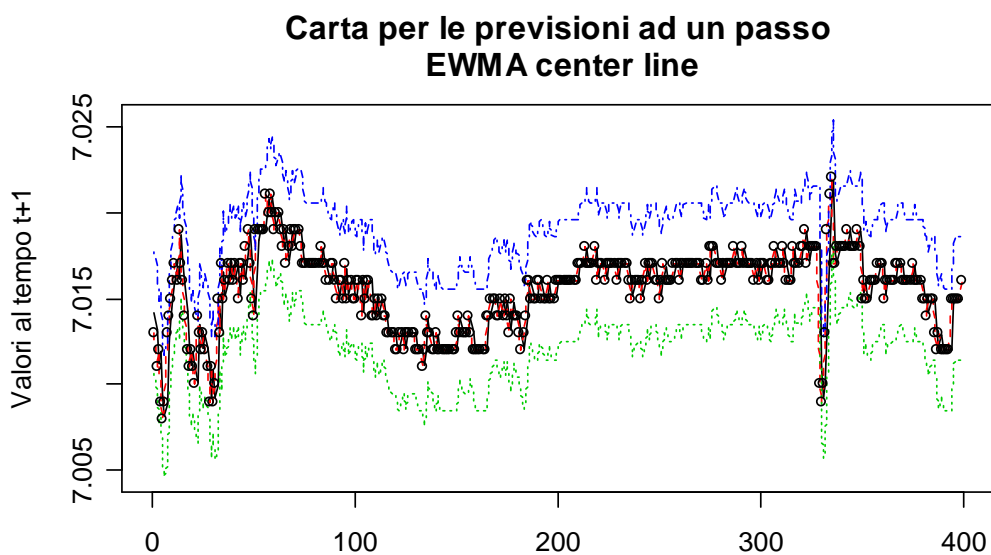


Figura 4.7: Carta per le previsioni ad un passo, Codice 755

Applichiamo anche una carta EWMA ed una Shewhart ai residui del modello di regressione adottato (Figure 4.8 e 4.9):

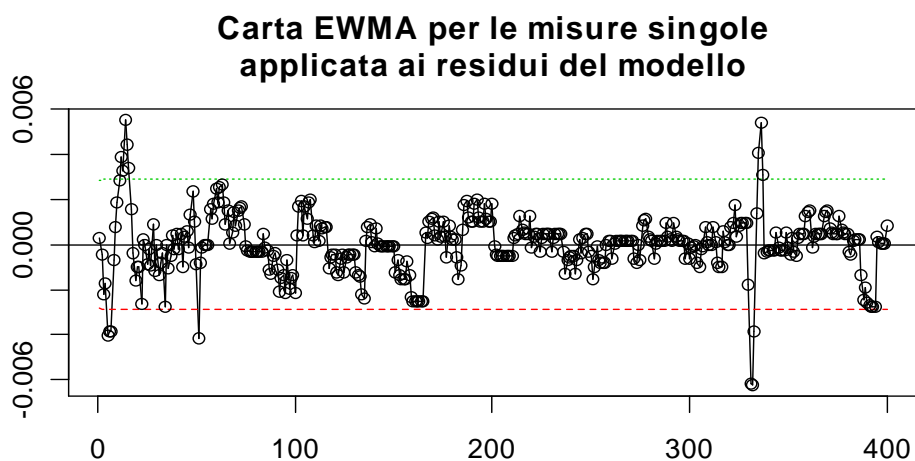


Figura 4.8: Carta EWMA applicata ai residui del modello adattato al Codice 755

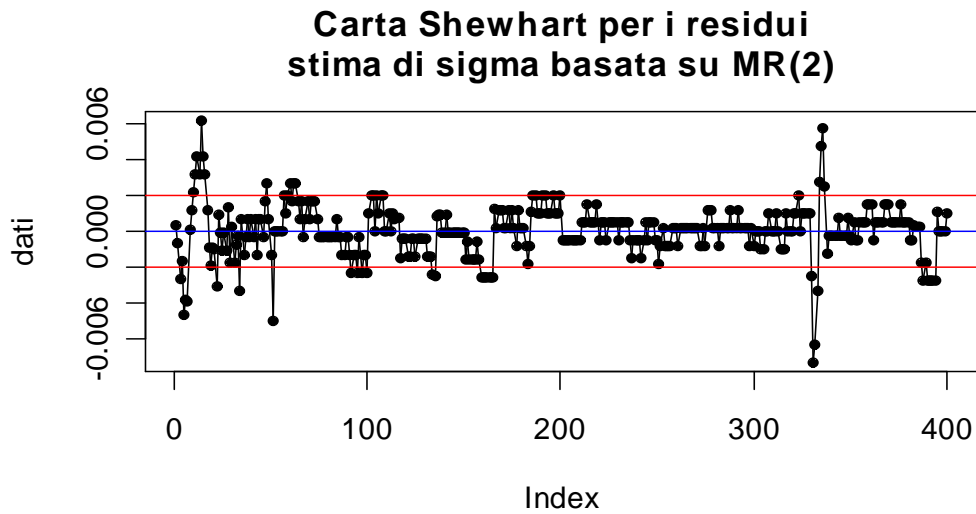


Figura 4.9: Carta Shewhart applicata ai residui del modello adattato al Codice 755

Notiamo la presenza di parecchi valori che escono dai limiti di controllo, tuttavia l'andamento dei punti è sufficientemente vicino alla linea centrale. Costruendo una carta EWMA *center line*, sempre applicata ai residui del modello, otteniamo il grafico in Figura 4.10:

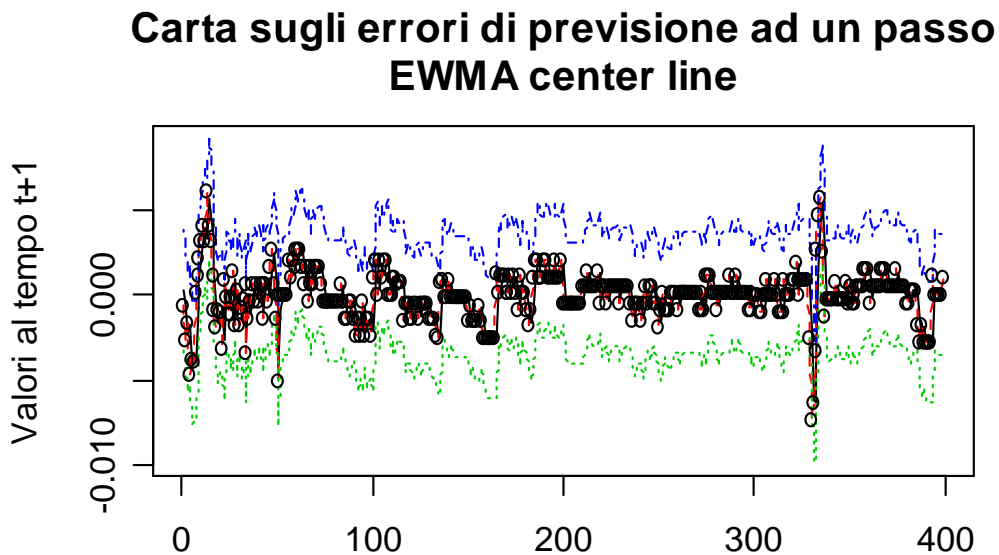


Figura 4.10: Carta EWMA center line applicata ai residui del modello del Codice 755

Il  $\lambda^{ott}$  vale 0.8 quindi otteniamo dei limiti che si basano sulle osservazioni vicine nel tempo.

In tutte le carte di controllo presentate vi sono delle osservazioni che cadono fuori dai limiti di specifica. L'osservazione da fare in questo caso è che tali dati sono sempre gli stessi, sia nello studio delle variabili originali sia per l'analisi dei residui. Essi sono presenti nella parte iniziale ed in corrispondenza dell'osservazione 320.

Ciò sta ad indicare che siamo in presenza di altre fonti di variabilità dovute a delle cause che non sono state prese in considerazione nello studio del modello (per esempio un possibile surriscaldamento della macchina).

## CAPITOLO 5

### **Capacità**

#### 5.1 *Indici di capacità*

Le Carte di Controllo sono un potente mezzo per mantenere un processo sotto controllo statistico, indicando le azioni correttive che devono essere intraprese al fine di eliminare le cause di variabilità indesiderata.

Tuttavia esse non tengono conto delle specifiche a cui il processo deve attenersi, come ad esempio le tolleranze di lavorazione o altre caratteristiche richieste. Il loro utilizzo non è dunque sufficiente per comprendere la reale capacità di un processo, né come questo può essere migliorato. A questo scopo vengono definiti gli indici di capacità, che rappresentano l'abilità del processo preso in esame di generare prodotti conformi alle specifiche. Essi, inoltre, permettono di riassumere in modo molto conciso i dati di un processo produttivo con il vantaggio di essere quantità adimensionali e quindi facilmente interpretabili e paragonabili tra loro.

Gli indici più comunemente usati sono il  $C_p$  e il  $C_{pk}$  ed essi richiedono che i dati seguano approssimativamente una distribuzione normale.

$C_p$  è l'indice che misura il "potenziale del processo". ed è definito come:

$$C_p = \frac{LTS - LTI}{6\sigma} \quad 5.1$$

dove  $LTS$  e  $LTI$  indicano rispettivamente il Limite di Tolleranza Superiore e il Limite di Tolleranza Inferiore e  $\sigma$  la deviazione standard del processo.

$C_p$  è dunque la misura del rapporto tra la dispersione ammissibile per il processo, calcolata per mezzo dalla differenza tra i Limiti di Tolleranza, e

la dispersione reale, rappresentata dal valore  $6\sigma$ , detta anche Tolleranza Naturale.

Generalmente per valutare tale indice si fa riferimento a dei valori critici tali per cui se:

- $C_p < 1$  : bassa capacità
- $1 < C_p < 1.6$  : media capacità
- $C_p \geq 1.6$  : alta capacità

$C_p$  indica dunque quanto un processo è "capace" solo se è centrato. Per questo motivo si introduce l'indice  $C_{pk}$ , che considera anche la posizione del processo rispetto ai Limiti di Tolleranza.

L'indice  $C_{pk}$  misura sia la dispersione che la centratura del processo, tenendo conto quindi dell'ampiezza della distribuzione e della posizione del livello medio del processo. Viene definito come:

$$C_{pk} = \frac{\min\left\{\frac{LTS - \mu}{\sigma}, \frac{\mu - LTI}{\sigma}\right\}}{3} \quad 5.2$$

Scegliendo il minore dei due valori calcolati, si determina quanto è capace il processo sul lato peggiore, quello cioè rappresentato dalla coda della gaussiana più vicina al limite di tolleranza.

Se:

- $C_{pk} > 1$  : l'ampiezza  $6\sigma$  dei dati cade completamente nei Limiti di Tolleranza;
- $C_{pk} = 1$  : indica che il 99,73% delle parti prodotte è nei limiti di tolleranza, cioè devono essere respinte solo 3 parti su 1000;
- $0 < C_{pk} < 1$  : indica che una parte dei prodotti del processo cade oltre i limiti;
- $C_{pk} = 0$  : indica che il processo è centrato su uno dei limiti di specifica;
- $C_{pk} < 0$  : indica che la media del processo giace al di

fuori limiti di specifica.

Per i calcoli applicati ai dataset abbiamo utilizzato un altro indice detto  $Pp$  definito come:

$$Pp = \frac{LTS - LTI}{6\sigma} \quad 5.3$$

Questo indice ha la stessa forma del  $Cp$ , ma differisce nel calcolo della deviazione standard. In questo caso  $\sigma$  viene posto pari alla deviazione standard complessiva (ovvero di lungo periodo).

## 5.2 Capacità calcolate sul Codice 653

Se vogliamo calcolare l'indice  $Cp$  per il Codice 653 ci dobbiamo trovare la deviazione standard del processo mediante lo scarto quadratico medio calcolato sulle escursioni mobili viste nel capitolo 4.1.1. Nel nostro caso l'indice risulta:

$$Cp = \frac{10.034 - 10.018}{6 \cdot 0.00043754} = 6.094681$$

Tale valore è molto più grande della soglia di accettabilità, quindi possiamo affermare che per quanto riguarda la dispersione il processo è buono.

Le specifiche bilaterali valgono:

$$LTS = 18.35261$$

$$LTI = -18.21548$$

il minimo risulta essere  $LTI$  e verrà quindi utilizzato nel calcolo del  $Cpk$ :

$$Cpk = \frac{LTI}{3} = 6.071826$$

Calcoliamo anche l'indice  $Pp$ :

$$Pp = \frac{10.034 - 10.018}{6 \cdot 0.001388835} = 1.920074$$

Tale indice assume un valore di molto inferiore al  $C_p$ , ciò sta ad indicare che il processo non è stabile.

Se facciamo un confronto mediante rapporto tra gli indici  $C_p$  e  $C_{pk}$

$$\frac{C_p}{C_{pk}} = 1.0037$$

vediamo che otteniamo praticamente l'unità. Questo ci indica che la media del processo è esattamente centrata sul punto medio della specifica.

Complessivamente possiamo affermare che il nostro processo è capace ma che non risulta stabile.

### 5.3 Capacità calcolate sul codice 755

L'indice  $C_p$  per il dataset del Codice 755 è dato da:

$$C_p = \frac{7.022 - 7.012}{6 \cdot 0.00054962} = 3.032398$$

Il valore ottenuto indica che il processo ha un'elevata capacità e ciò ci lascia un po' dubbiosi in quanto, nelle analisi effettuate precedentemente, esso sembrava molto variabile.

Le specifiche bilaterali per il calcolo dell'indice  $C_{pk}$  valgono:

$$LTS = 11.71264$$

$$LTI = -6.481751$$

dalle quali risulta che il minimo, sempre in valore assoluto, corrisponde a  $LTI$  quindi il  $C_{pk}$  sarà calcolato mediante:

$$C_{pk} = \frac{LTI}{3} = 2.160584$$

L'indice  $P_p$  vale:

$$P_p = \frac{7.022 - 7.012}{6 \cdot 0.00230203} = 0.7239987$$



Sostanzialmente abbiamo dei valori di  $C_p$  e di  $C_{pk}$  buoni, in quanto la varianza a media mobile nei dati risulta contenuta e possiamo affermare che il processo è sostanzialmente in controllo.

Infine un valore di  $P_p$  molto inferiore ai precedenti indici, in quanto la varianza totale nei dati risulta essere elevata.

Tuttavia avevamo visto che i dati non si distribuivano normalmente quindi non possiamo fare queste affermazioni con sicurezza. Infatti si era detto che uno degli assunti principali per il calcolo degli indici di capacità consiste nel fatto di avere dati provenienti da una distribuzione approssimativamente normale.



## CONCLUSIONI

Lo scopo di questo elaborato era quello di analizzare l'andamento di due processi impiegati nella lavorazione di ingranaggi.

Sono partita effettuando un'analisi esplorativa sui dati raccolti nel periodo di stage svolto presso l'azienda mG miniGears. Per realizzare tale analisi ho esplorato la variabilità complessiva ed entro i sottogruppi. Ho utilizzato degli istogrammi e dei boxplot, per avere una prima idea sulla distribuzione dei dati e successivamente, per verificare se si distribuivano secondo una legge normale ho costruito dei *normal probability plot* e applicato delle statistiche test. Per vedere se le osservazioni fossero o meno correlate ho disegnato i grafici dell'autocorrelazione totale e parziale ed infine dei grafici di autodispersione.

Ho adattato dei modelli lineari ai dati per verificare se nei processi interviene della variabilità di tipo non casuale; per questi modelli ho calcolato i residui analizzandoli mediante istogrammi, diagrammi di dispersione e *normal probability plot*.

Infine, per giungere all'obiettivo della tesi, ho applicato delle carte di controllo ai dati per verificare se il processo sia o meno in uno stato di controllo statistico e calcolato alcuni indici di capacità.

Sono state riportate anche delle carte di controllo calcolate sui residui dei modelli adattati ai due ingranaggi, in quanto si è visto che non abbiamo la presenza della sola variabilità naturale, ma esiste l'effetto significativo di alcune variabili sulla variabilità del diametro dei fori.

Per quanto riguarda il Codice 653 possiamo affermare che il processo sembra stabile.

Abbiamo visto che la media complessiva delle osservazioni non si discosta dal valore nominale fissato nelle specifiche di progetto, mentre le medie calcolate nei campioni risultano differenti dal valore target.

Dato che i dati sono fortemente autocorrelati abbiamo applicato una carta EWMA direttamente sui dati originali. La carta di controllo ha evidenziato le differenze in media tra i gruppi.

Il modello di regressione adattato ai dati è risultato molto apprezzabile, tuttavia possiamo dire che i residui non sono completamente soddisfacenti. L'introduzione di altre variabili esplicative potrebbe portare ad un modello che meglio interpreta i dati.

Comunque la capacità del processo di produrre pezzi conformi alle specifiche è provata dai valori elevati degli indici di capacità.

Essendo questo un codice critico possiamo affermare che le procedure utilizzate dagli operatori sono adeguate per soddisfare i clienti, che generalmente richiedono un  $Cpk$  di 2.

Per il Codice 755 siamo in presenza di osservazioni molto variabili e che addirittura escono dai limiti di tolleranza all'inizio e verso l'osservazione 320.

Con gli istogrammi e i boxplot applicati ai valori abbiamo scartato l'ipotesi che il campione si distribuisca come una variabile casuale normale.

Alle osservazioni, che risultano essere autocorrelate, è stato applicato un modello che non coglie tutta la variabilità presente. Questo fatto si riscontra nelle carte di controllo, dove alcuni punti cadono fuori dai limiti. Tali punti si ripetono nelle carte applicate ai residui e ciò sta ad indicare la possibile presenza di un'ulteriore variabile da inserire nel modello. I residui calcolati, infatti, non sono pienamente soddisfacenti.

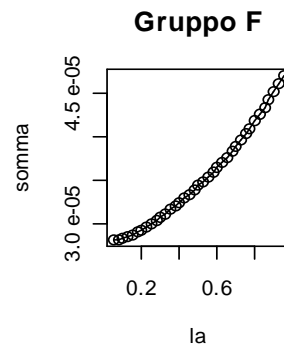
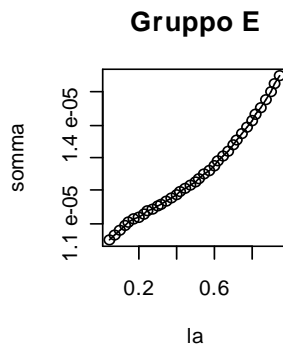
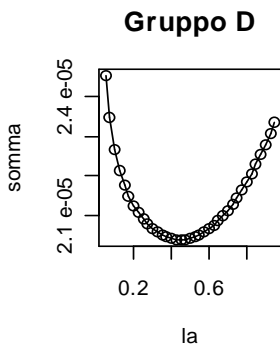
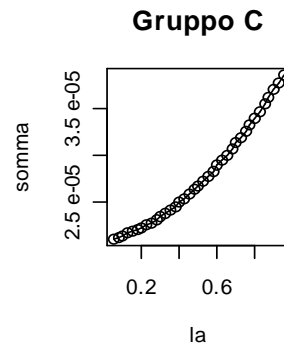
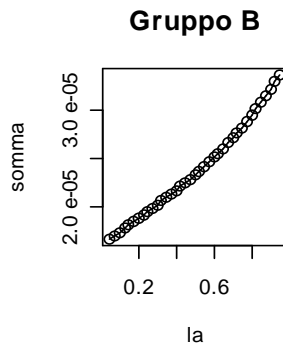
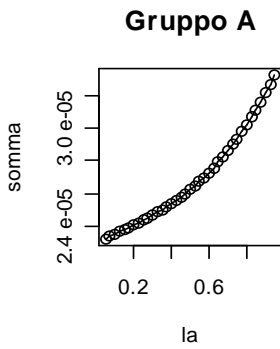
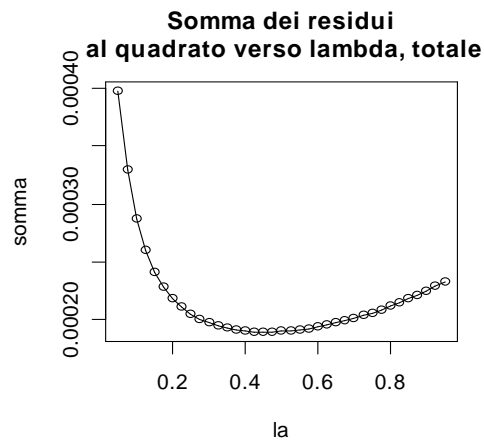
Gli indici  $Cp$  e  $Cpk$  ci dicono che il processo in termini di controllo è accettabile, seppur con valori meno elevati di quelli ottenuti per il Codice 653.

Tuttavia l'indice di *performance*  $Pp$  fa notare che una elevata variabilità di lungo periodo e di conseguenza una parziale instabilità del processo produttivo. Bisognerebbe perciò ricorrere ad un'analisi più approfondita per individuare ed eliminare le principali cause di variabilità.

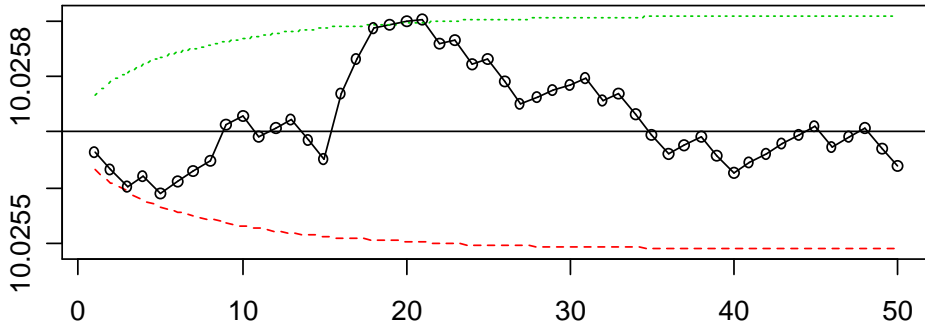
# APPENDICE

A: I grafici

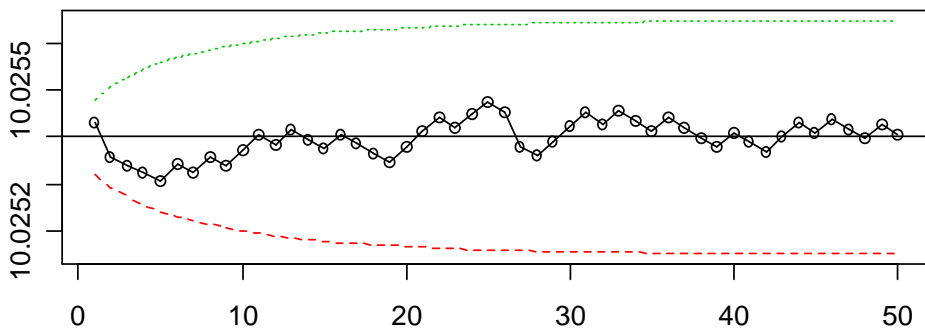
Codice 653:



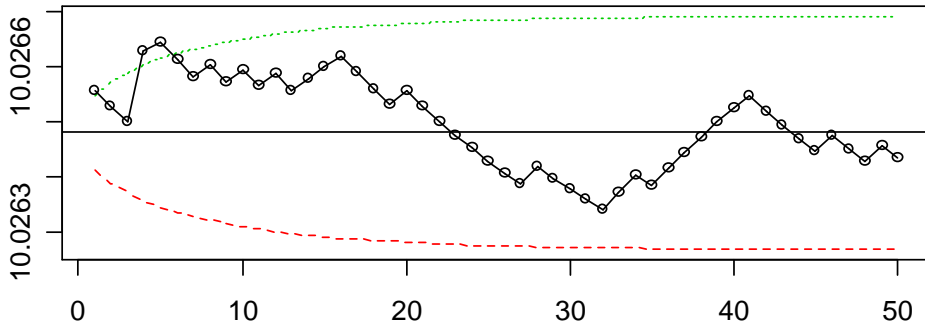
**Carta EWMA per le misure singole, lambda = 0.05 Gruppo A**



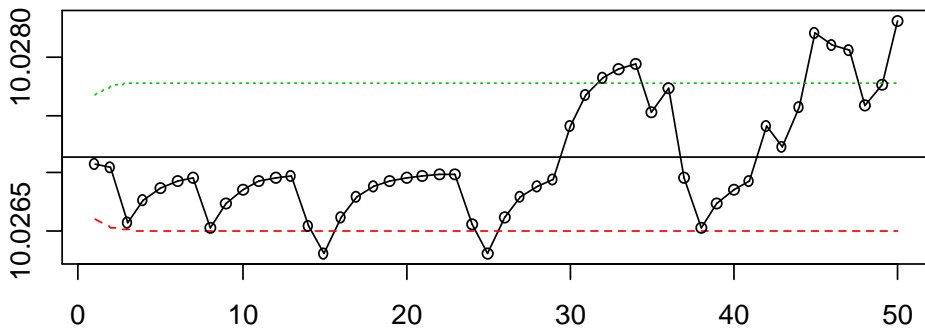
**Carta EWMA per le misure singole, lambda = 0.05 Gruppo B**



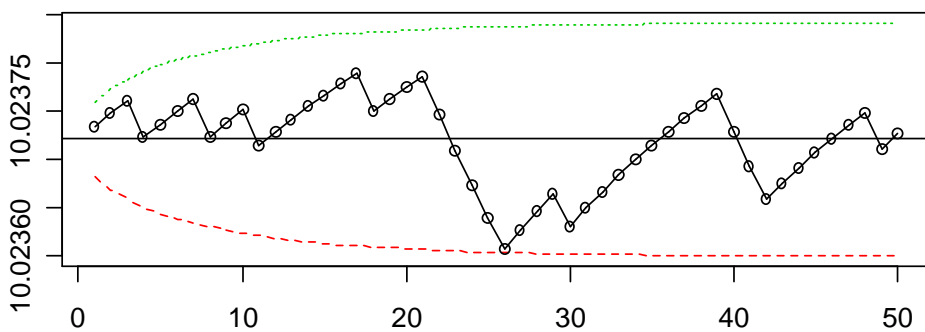
**Carta EWMA per le misure singole, lambda = 0.05 Gruppo C**



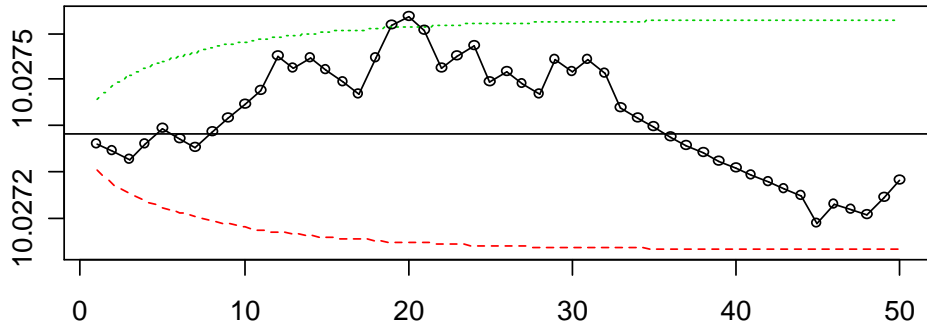
**Carta EWMA per le misure singole, lambda = 0.45 Gruppo D**



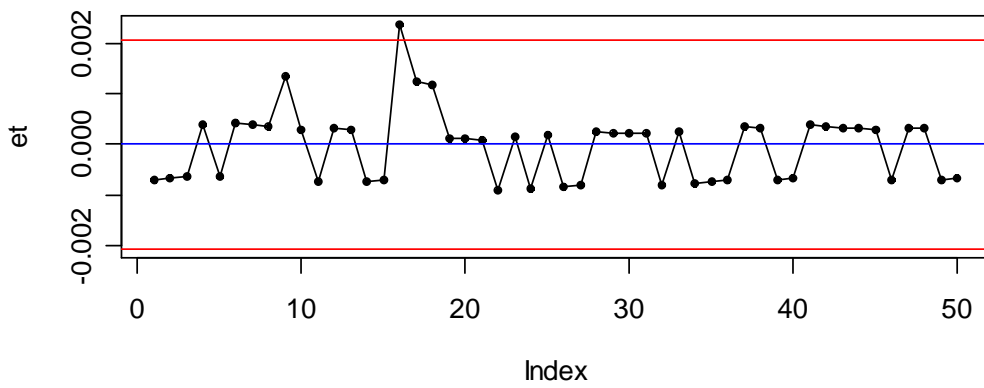
**Carta EWMA per le misure singole, lambda = 0.05 Gruppo E**



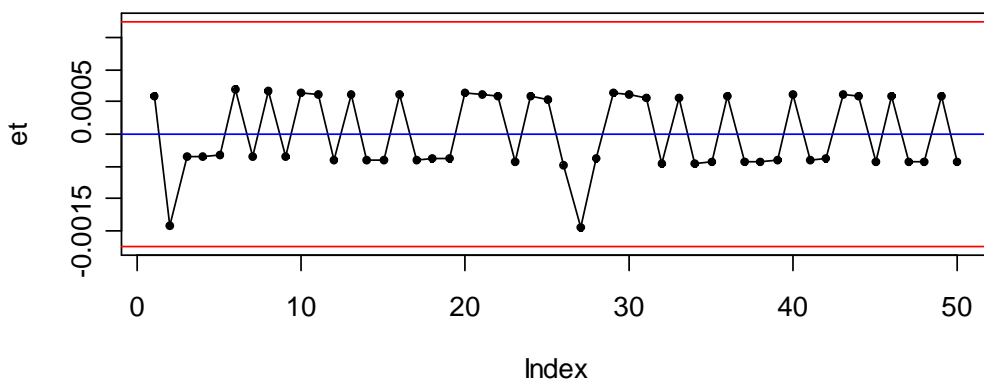
**Carta EWMA per le misure singole, lambda = 0.05 Gruppo F**



**Carta sugli errori di previsione ad un passo Gruppo A**

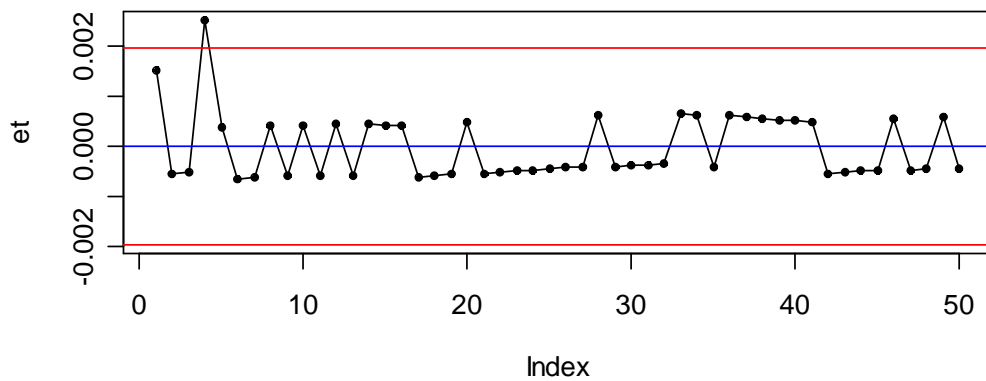


**Carta sugli errori di previsione ad un passo Gruppo B**

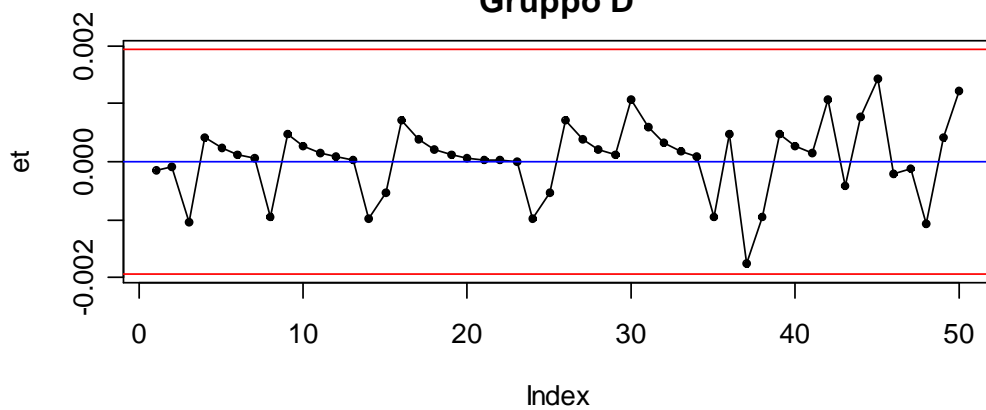




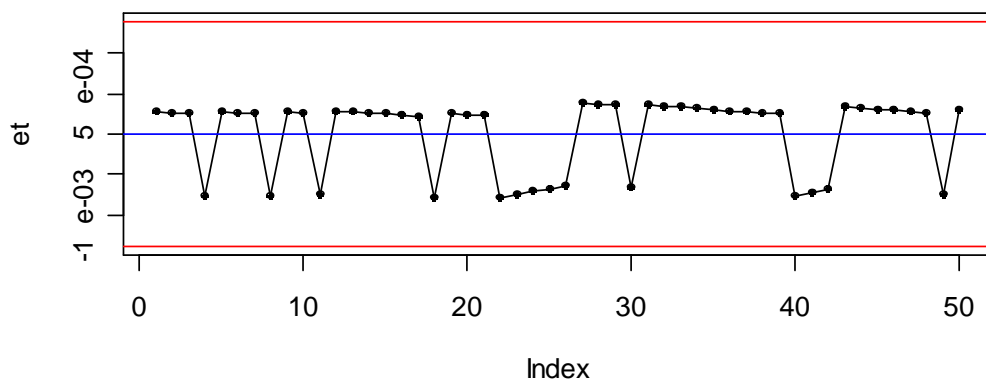
**Carta sugli errori di previsione ad un passo  
Gruppo C**



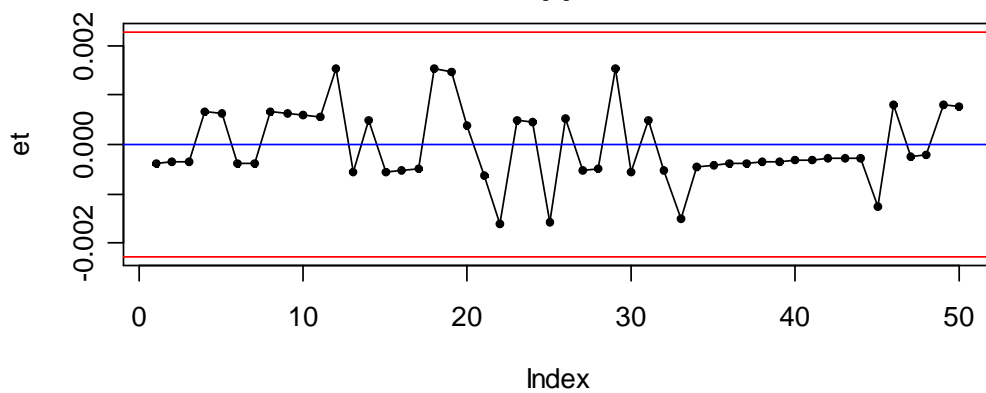
**Carta sugli errori di previsione ad un passo  
Gruppo D**



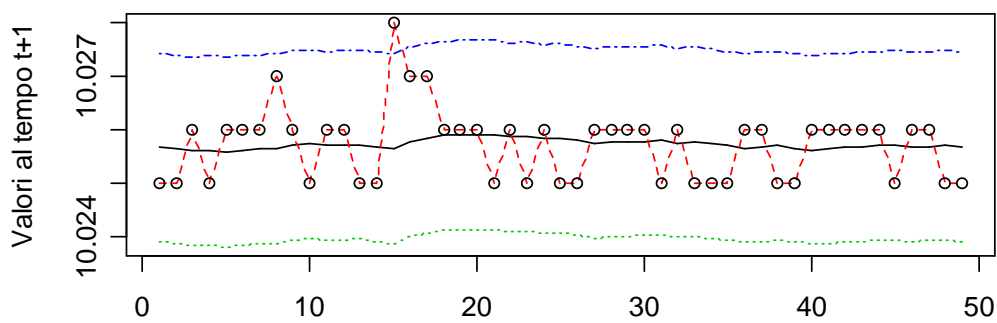
**Carta sugli errori di previsione ad un passo  
Gruppo E**



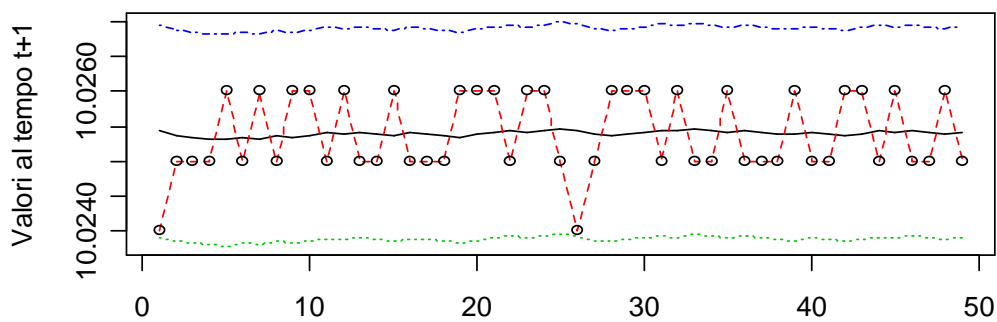
**Carta sugli errori di previsione ad un passo  
Gruppo F**



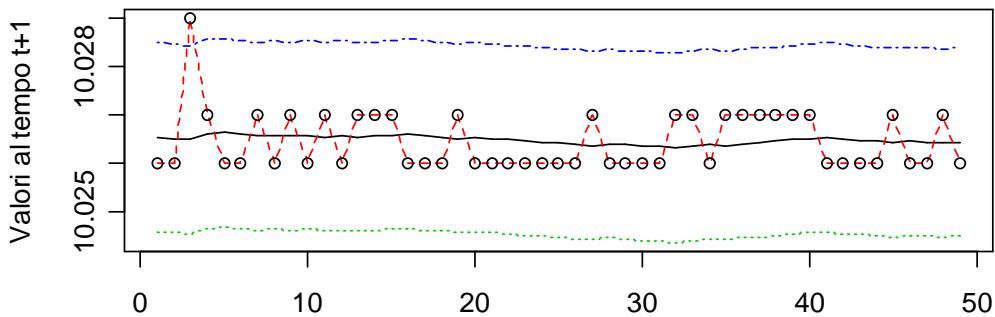
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo A**



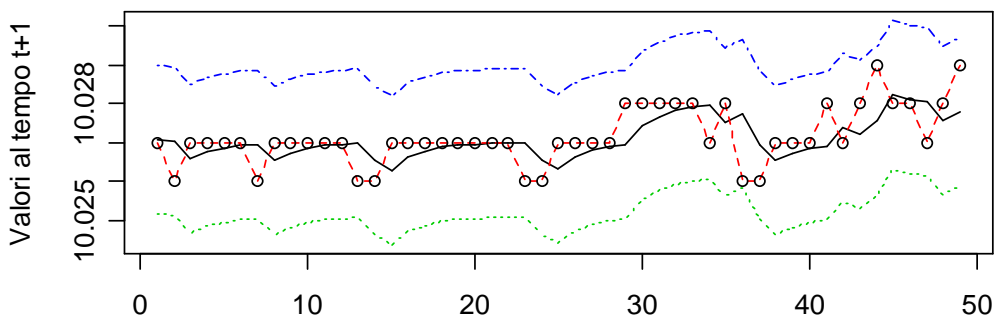
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo B**



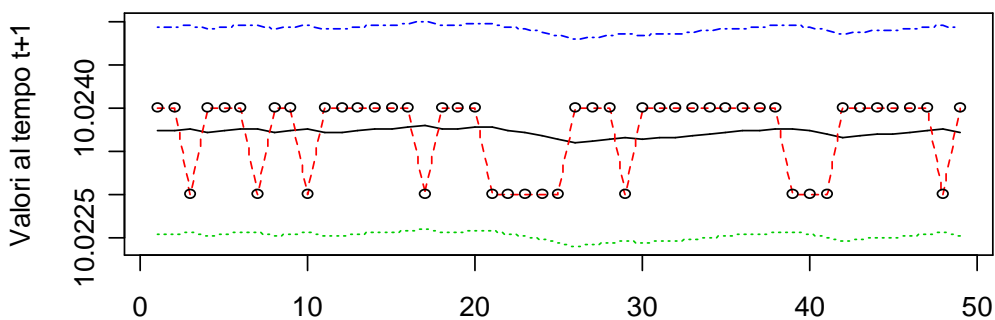
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo C**



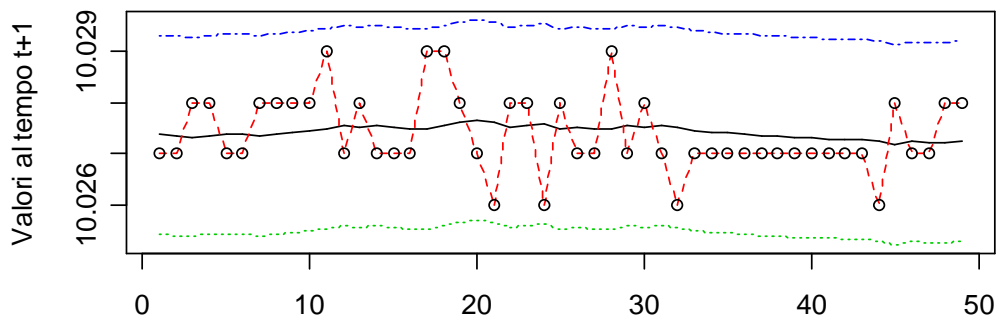
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo D**



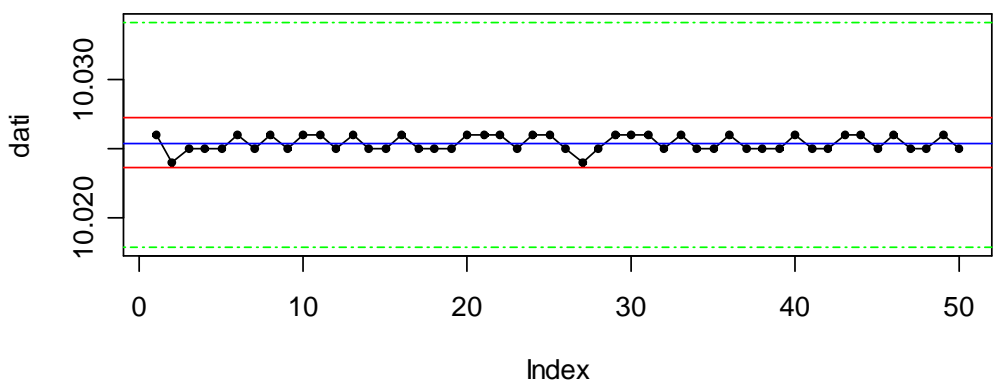
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo E**



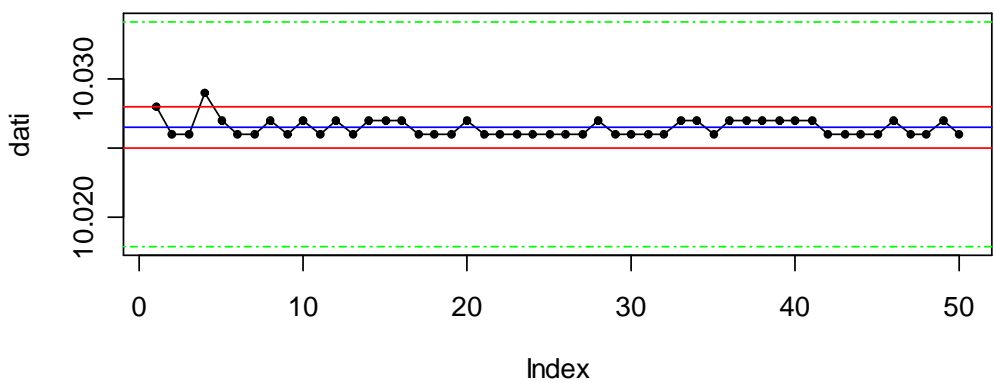
**Carta per le previsioni ad un passo  
EWMA center line - Gruppo F**



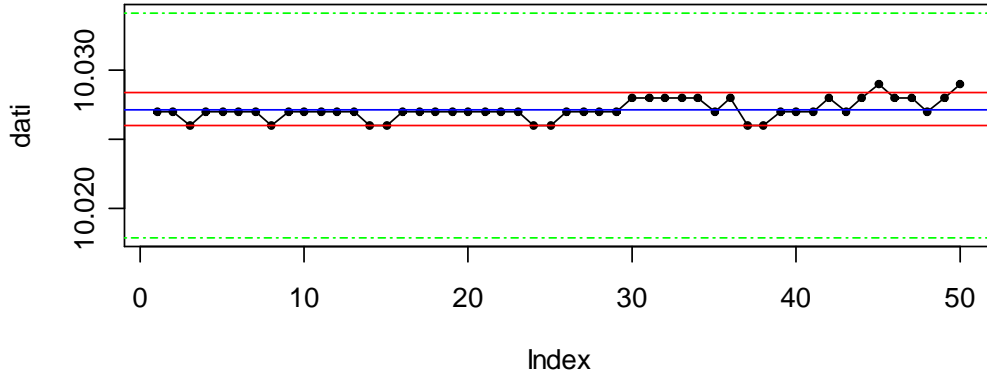
**Carta Shewart per le misure singole  
stima di sigma basata su MR(2) Gruppo B**



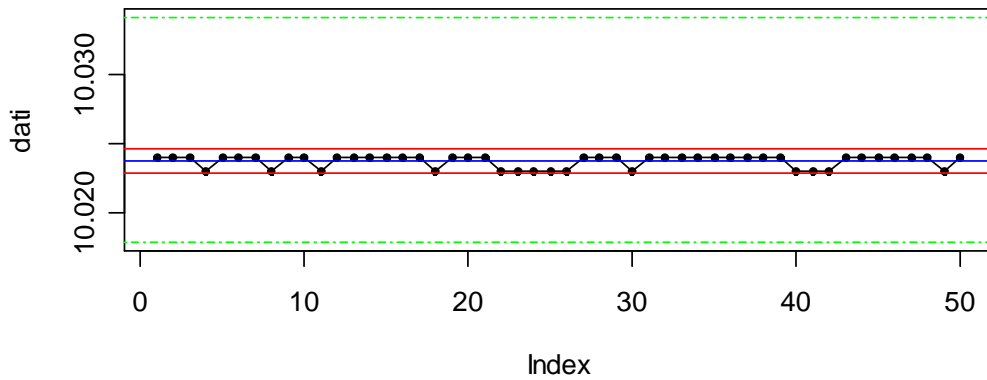
**Carta Shewart per le misure singole  
stima di sigma basata su MR(2) Gruppo C**



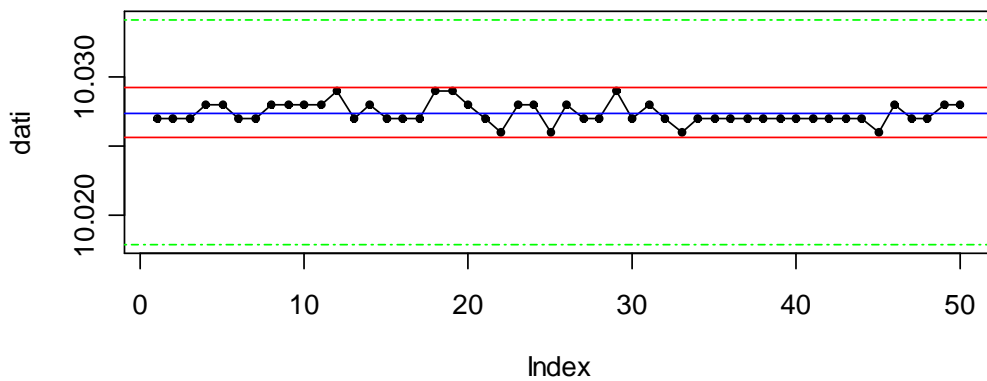
**Carta Shewart per le misure singole  
stima di sigma basata su MR(2) Gruppo D**



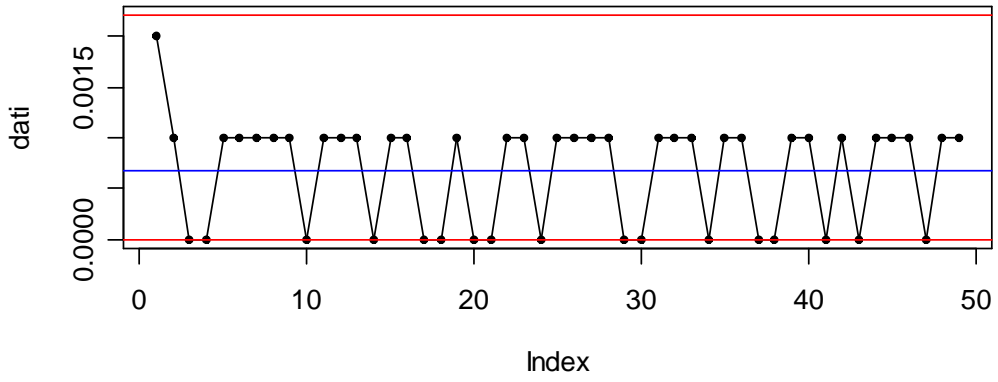
**Carta Shewart per le misure singole  
stima di sigma basata su MR(2) Gruppo E**



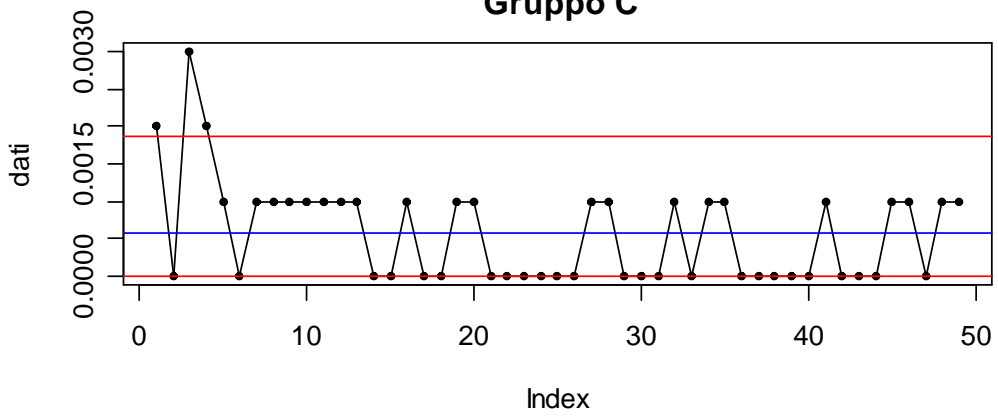
**Carta Shewart per le misure singole  
stima di sigma basata su MR(2) Gruppo F**



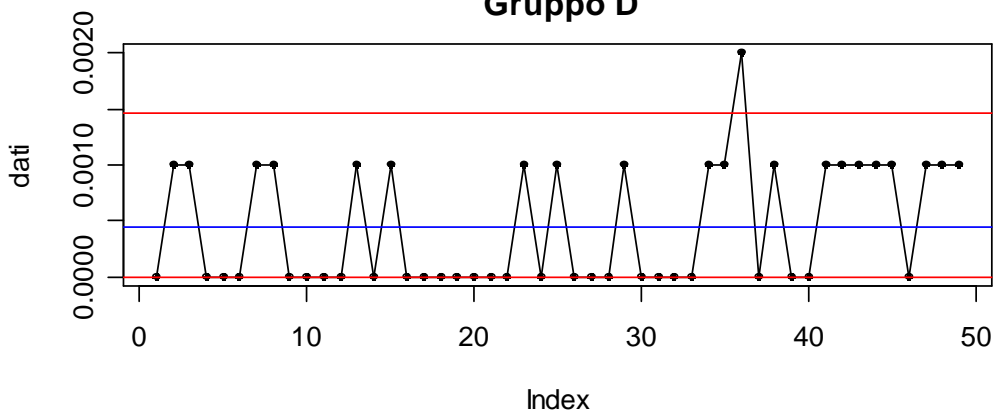
**Carta sulle escursioni mobili  
Gruppo B**



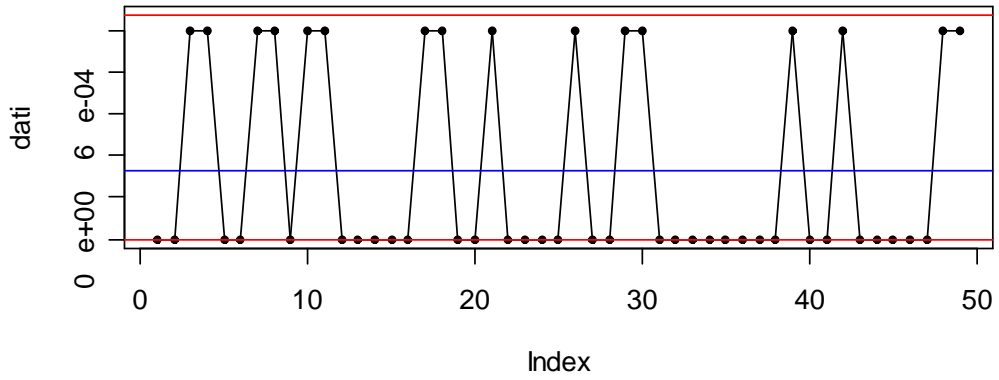
**Carta sulle escursioni mobili  
Gruppo C**



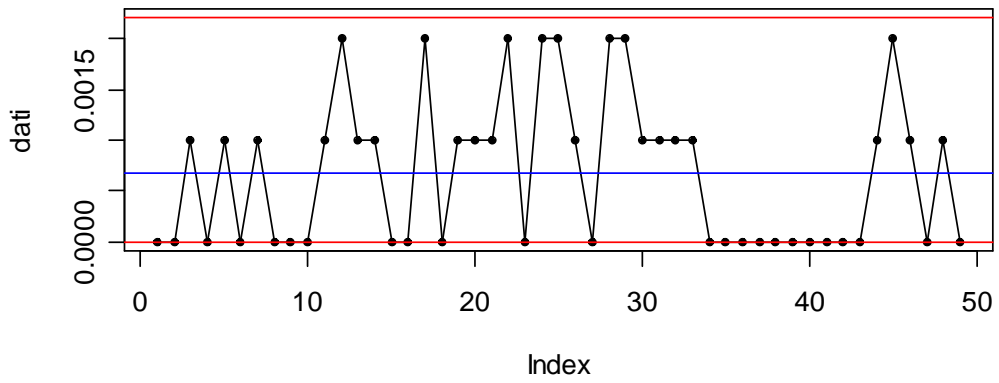
**Carta sulle escursioni mobili  
Gruppo D**



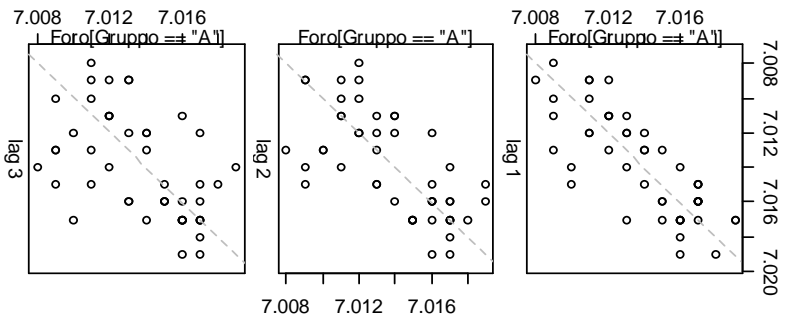
**Carta sulle escursioni mobili  
Gruppo E**



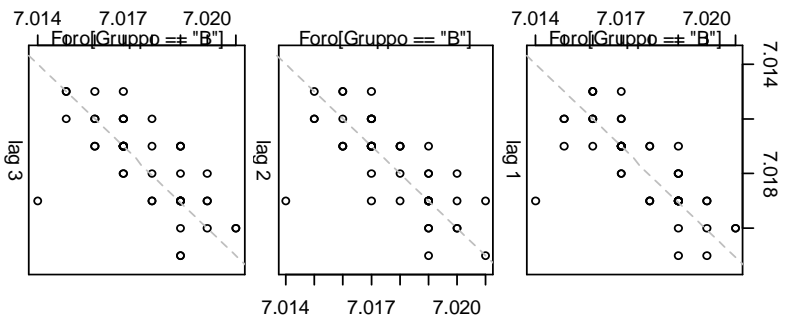
**Carta sulle escursioni mobili  
Gruppo F**



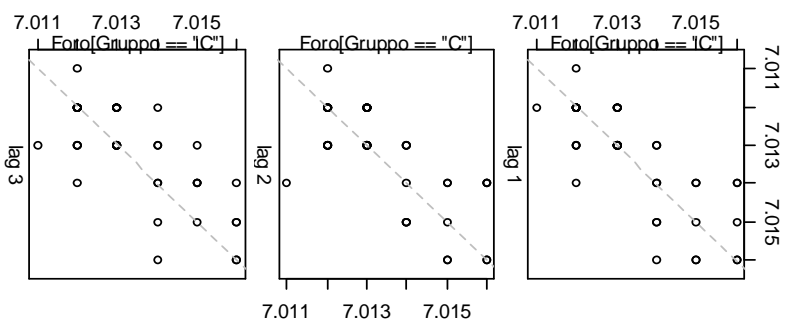
**Gruppo A**



**Gruppo B**

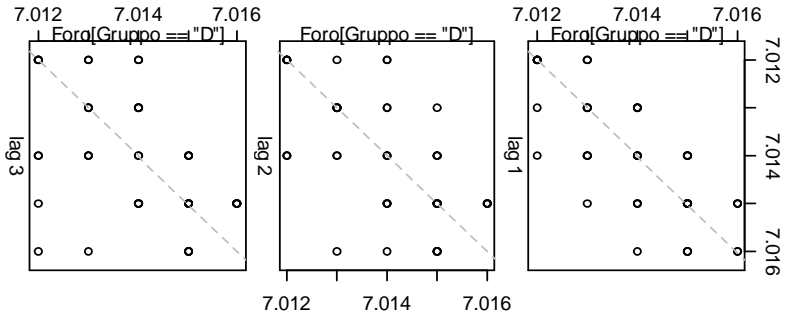


**Gruppo C**

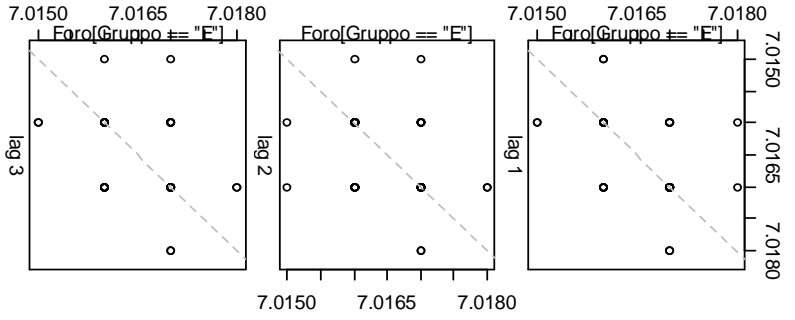




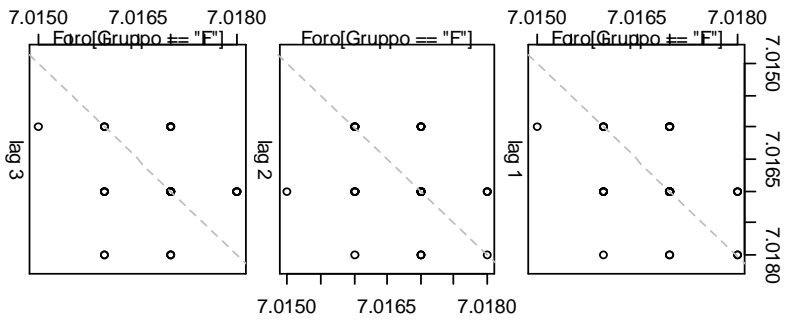
**Gruppo D**



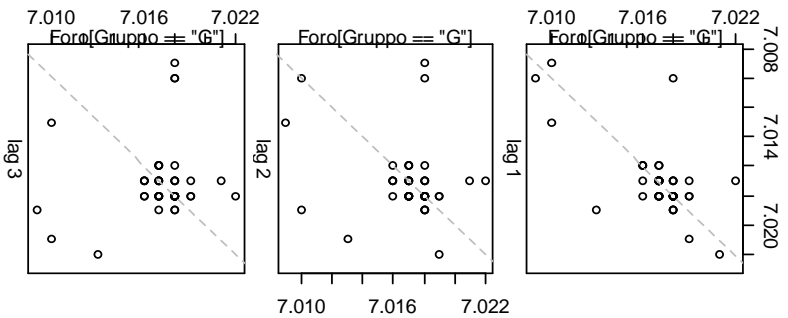
**Gruppo E**



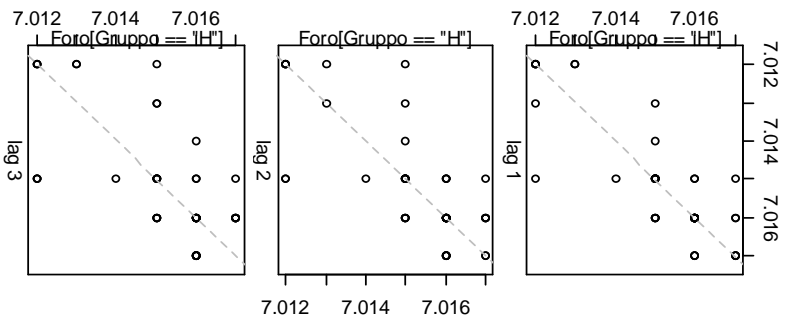
**Gruppo F**



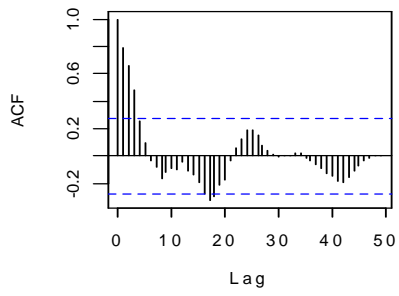
### Gruppo G



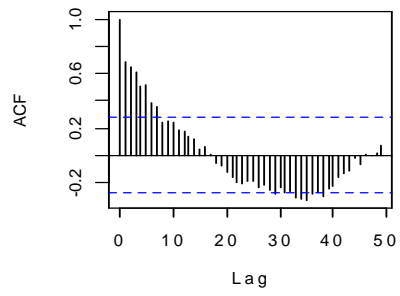
### Gruppo H



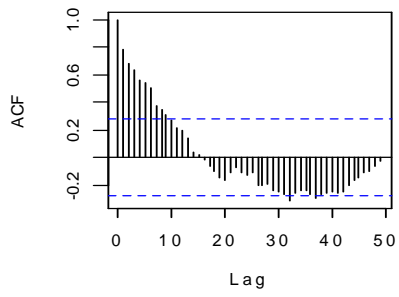
**Gruppo A**



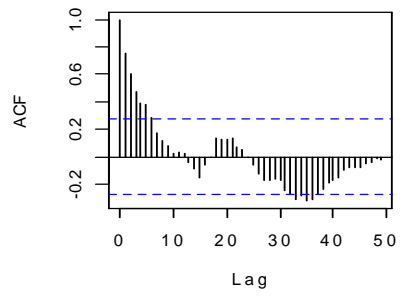
**Gruppo B**



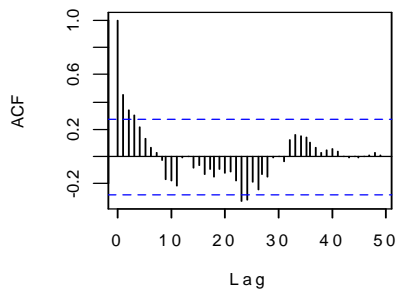
**Gruppo C**



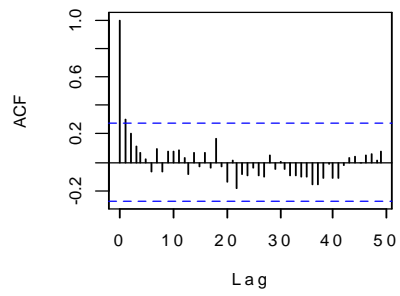
**Gruppo D**



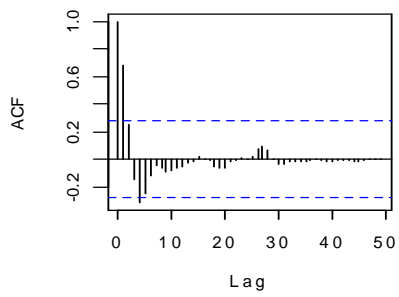
**Gruppo E**



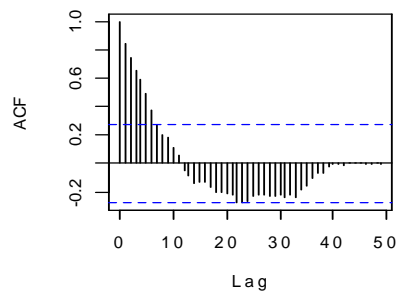
**Gruppo F**



**Gruppo G**



**Gruppo H**



## B:Le funzioni utilizzate

- Per leggere le osservazioni singole:

```
wsing<-function(dati)
  {MR2 <- abs(diff(dati))
  MR2}
```

- Per stimare il valore atteso e la varianza mediante le escursioni mobili

```
vaMR2 <- function(dati)
  {wi <- wsing(dati)
  MRmedio <- mean(wi)
  cat("MRmedio = ", round(MRmedio, digits = 8), "\n")
  d2<-c(1.128,1.693,2.059,2.326,2.534,2.70,2.847,2.970,3.078)
d3<-c(0.853,0.888,0.880,0.864,0.848,0.833,0.820,0.808,0.80)
  stimaMR <- MRmedio / d2[1]
  cat("stimaMR = ", round(stimaMR, digits = 8), "\n")
  vaMR <- d2[1] *
  cat("vaMR = ", round(vaMR, digits = 8), "\n")
  sigmaMR <- stimaMR * d3[1]
  cat("sigmaMR = ", round(sigmaMR, digits = 8), "\n")
  risMR <- round(c(MRmedio, stimaMR, vaMR, sigmaMR), digits =
8)}
```

- Carta per il controllo della media

```
Shms <- function(dati, L, muw, sqmw, tit, tolinf, tolsup)
  {LCL <- muw - (L*sqmw)
  LC <- muw
  UCL <- muw + (L*sqmw)
  cat("LCL = ", round(LCL, digits = 8), "\n")
  cat("LC = ", round(LC, digits = 8), "\n")
  cat("UCL = ", round(UCL, digits = 8), "\n")
  plot(dati,ylim = c(tolinf, tolsup), main = tit, type = "l")
  points(dati, pch = 20)
  abline(a = LCL, b = 0, col = "red")
  abline(a = LC, b = 0, col = "blue")
  abline(a = UCL, b = 0, col = "red")
  abline(a = tolinf, b = 0, col = "green", lty = 4)
  abline(a = tolsup, b = 0, col = "green", lty = 4)}
```

- Carta per il controllo della varianza

```
Shvar <- function(dati, L, muw, sqmw, tit)
  {LCL <- muw - L*sqmw
  if (LCL < 0) LCL<-0
  LC <- muw
  UCL <- muw + L*sqmw
  plot(dati, ylim = c(LCL, max(UCL, dati)), main = tit, type =
"l")
  cat("LCL = ", round(LCL, digits = 8), "\n")
  cat("LC = ", round(LC, digits = 8), "\n")
  cat("UCL = ", round(UCL, digits = 8), "\n")
  points(dati, pch = 20)
```

```

abline(a = LCL, b=0, col = "red")
abline(a = LC, b=0, col = "blue")
abline(a = UCL, b=0, col = "red")}

```

- Per calcolare il lambda ottimale e la stima di  $\sigma_p$

```

graf1 <- function(dati, w0)
{la <- seq(0.05, 0.95, by = 0.025)
x <- 1:length(la)
somma <- numeric(length(la))
y <- matrix(0, ncol = length(la), nrow = length(dati))
for (i in x)
  {y[,i] <- ewmae(dati, la[i], w0)[,2]
  somma[i] <- sum(y[,i])}
i <- i + 1
plot(la, somma, type = "l", main = "Somma dei residui\n al
quadrato verso lambda")
points(la, somma)
ris <- cbind(la, somma)
j <- min(ris[,2])
lott <- ris[ris[,2]==j, 1]
cat("lambda ottimale = ", lott, "\n")
cat("somma dei residui al quadrato = ", j, "\n")
cat("sigma2p = ", j / length(dati), "\n")
cat("sigmap = ", sqrt(j / length(dati)), "\n")}

```

- Calcola le previsioni ad un passo

```

ewmae <- function(dati, la, w0)
{k <- length(dati)
x <- c(rep(0, k))
e <- c(rep(0, k))
for (t in 1:length(dati))
  {if (t==1)
    {x[1] <- w0 + la * (dati[1] - w0)
    e[1] <- (dati[1] - w0)}
  else
    {x[t] = x[t-1] + la * (dati[t] - x[t-1]);
    e[t] <- (dati[t] - x[t-1])}
  }
t <- t + 1
ris <- cbind(e, e^2)
ris}

```

- Calcola i limiti della Carta EWMA

```

limiti <- function(dati, la, L, media, sd, n)
{k <- length(dati)
sqm <- sd / sqrt(n)
lst <- c(rep(0, k))
lit <- c(rep(0, k))
for (t in 1:k)
  {lit[t] <- media - (L * sqm * sqrt((la / (2 - la)) *
(1 - (1 - la)^(2*t))))
  lst[t] <- media + (L * sqm * sqrt((la / (2 - la)) * (1
- (1 - la)^(2*t))))}

```

```

t <- t + 1
li <- media - (L * sqm * sqrt((la / (2 - la))))
ls <- media + (L * sqm * sqrt((la / (2 - la))))
rbind(cbind(lit, lst), c(li, ls))}

```

- Disegna la carta EWMA

```

grafEWMA <- function(dati, la, L, media, sd, n, tit, ylab)
{wewma <- ewma(dati, la, media)
lim <- limiti(dati, la, L, media, sd, n)
lim2 <- lim[1:length(dati),]
matplot(cbind(wewma, lim2), type = "l", lty = 1:3, main =
tit, ylab=ylab)
points(wewma)
abline(a = media, b = 0)}

```

- Calcola i valori della carta EWMA

```

ewma <- function(dati, la, media)
{k <- length(dati)
x <- c(rep(0, k))
for (t in 1:length(dati))
{if (t==1)
x[1] <- la * dati[1] + (1 - la) * media
else
x[t] = la * dati[t] + (1 - la) * x[t-1]}
t <- t + 1
x}

```

- Disegna la carta Shewhart per gli errori di previsione ad un passo

```

Grafet <- function(dati, lott, sigmap, w0, tit)
{LCL <- -3 * sigmap
LC <- 0
UCL <- 3 * sigmap
cat("LCL = ", round(LCL, digits = 8), "\n")
cat("LC = ", round(LC, digits = 8), "\n")
cat("UCL = ", round(UCL, digits = 8), "\n")
et <- ewmae(dati, lott, w0)[,1]
plot(et, ylim = c(min(LCL,et), max(UCL,et)), main = tit,
type = "l")
points(et, pch = 20)
abline(a = LCL, b = 0, col = "red")
abline(a = LC, b = 0, col = "blue")
abline(a = UCL, b = 0, col = "red")
et}

```

- Disegna la carta EWMA center line

```

grafEWMAc <- function(dati, lott, L, sigmap, w0, tit)
{w <- ewma(dati, lott, w0)
limc <- lim(dati, lott, L, sigmap, w0)
k <- length(dati) - 1
matplot(cbind(w[1:k],dati[2:length(dati)],limc[1:k,]), type
= "l", lty = 1:4, ylab = "Valori al tempo t+1", main = tit)
points(dati[2:length(dati)])}

```

- Calcola I limiti della carta EWMA center line

```
lim <- function(dati, lott, L, sigmap, w0)
  {k <- length(dati)
  lst <- c(rep(0, k))
  lit <- c(rep(0, k))
  w <- ewma(dati, lott, w0)
  for (t in 1:k)
    {lit[t] <- w[t] - L * sigmap
    lst[t] <- w[t] + L * sigmap}
  t <- t + 1
  cbind(lit, lst)}
```





## RIFERIMENTI BIBLIOGRAFICI

- MONTGOMERY D.C. (2000), “Il controllo statistico della qualità”, McGraw-Hill, 1999.
- IACUS S.M., MASAROTTO G., “Laboratorio di Statistica con R”, McGraw-Hill, 2003.
- Shewhart, WA (1931), “Economic control of quality of manufactured product”, D. Van Nostrand Company, Inc.
- “MINITAB Manuale Introduttivo”, GMSL S.r.l.2003
- Manuali in linea per il programma R
- Dispense delle lezioni
- [www.minigears.it](http://www.minigears.it)