

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN STATISTICA E TECNOLOGIE INFORMATICHE

TESI DI LAUREA

CONTROLLO STATISTICO DELLA QUALITA' PER DATI AUTOCORRELATI
E NON-NORMALI: IL CASO RINO MASTROTTO GROUP

Relatore: Ch.ma Prof.ssa GIOVANNA CAPIZZI

Laureanda: Angela Dal Lago

Matricola: 502550

ANNO ACCADEMICO 2005-2006

INDICE

CAPITOLO 1	5
1.1. Storia dell'azienda	5
1.2. Lo stage	6
1.3. Presentazione della tesi	7
CAPITOLO 2	9
2. Analisi dello spessore delle pelli	9
2.1. Analisi esplorativa dei dati	9
2.2. Introduzione alle carte di controllo	14
2.2.1. La carta Shewhart	15
2.2.2. La carta EWMA	16
2.2.3. Disegno della carta di controllo EWMA	18
2.2.4. Carta EWMA come previsore del livello medio del processo	19
2.2.5. Disegno di un'EWMA per dati autocorrelati	19
2.3. Applicazione delle carte di controllo ai dati	21
CAPITOLO 3	24
3.1. Carte di controllo basate sui ranghi	24
3.2. Ranghi standardizzati	24
3.3. Exponentially Weighted Moving Averages	25
3.4. Applicazione della carta EWMA basata sui ranghi	26
3.4.1. Analisi preliminare dei dati	27
3.4.2. Carta EWMA basata sui ranghi	28
CAPITOLO 4	31
4. Analisi della capacità del processo	31
4.1. Gli indici di capacità	31
4.2. Indici di capacità per dati non-normali	34
4.3. Indici di capacità sui dati osservati	36

CAPITOLO 5

38

5. Conclusioni

38

CAPITOLO 1

INTRODUZIONE

1.1 STORIA DELL'AZIENDA

Le origini storiche del RINO **MASTROTTO** GROUP risalgono al 1957 e, sin d'allora, esso si è caratterizzato per la sua notevole flessibilità in termini di struttura societaria: delle due aziende presenti inizialmente, si è arrivati oggi ad una struttura societaria complessa ed integrata.

Nel corso degli anni le aziende sono aumentate nel numero e cresciute nelle dimensioni, andando a costituire una delle maggiori realtà industriali del settore conciario nazionale e dell'area geografica di riferimento.

L'alto grado di automazione e l'impiego di tecnologie all'avanguardia consentono di utilizzare le risorse umane per le attività non standardizzabili, quali il controllo delle pelli e la scelta, che rivestono notevole importanza nella qualità del prodotto finale.

Il gruppo strategico costituitosi negli anni rappresenta ora una realtà chiara e ben definita di cui la RINO **MASTROTTO** GROUP S.p.A. è la capogruppo.

Negli ultimi anni la realtà del gruppo si è allargata con nuove aziende: la BERMAS LTDA situata in Brasile e CALBE SUD denominata dal 2004 APELLE S.p.A. nata dall'acquisto di un progetto industriale per la realizzazione nella Zona Industriale della Val Basento (Basilicata) di uno stabilimento per la rifinizione di pellame destinato prevalentemente al mercato locale del salotto imbottito.

RINO **MASTROTTO** GROUP S.p.A. nata nel 1998 dalla fusione di 6 aziende, costituisce oggi il fulcro del gruppo e centralizza presso di sé le funzioni ed i servizi chiave.

Con un sistema produttivo ed una struttura commerciale articolati in Divisioni (AREA FASHION, BASMAR, BRUSAROSCO, CALBE, GALASSIA e POMARI) ed un organico di circa 400 persone, la società produce oggi oltre 10 milioni di m² di pelle finita l'anno.

1.2 LO STAGE

Tra le varie Divisioni presenti nella RINO MASTROTTO GROUP S.p.A., la mia domanda di stage è stata accolta dalla Divisione CALBE – BRUSAROSCO sita in Zona Industriale ad Arzignano (VI).

La produzione che viene svolta in questa Divisione riguarda in particolar modo l'arredamento, la carrozzeria e la calzatura (soprattutto crosta cioè conciato che si ottiene dallo strato inferiore della pelle mediante la fase di spaccatura)

L'organizzazione esegue le attività di verifica ispettiva interne o esterne, per conto della direzione ed in accordo con quanto stabilito dalla direzione, al fine di dare evidenza che quanto pianificato sia mantenuto e risultati efficace.

Il piano di verifica comprende:

- le verifiche del sistema di gestione per la qualità in conformità alla ISO/TS 16949/2002 e alla ISO 9001:2000 per verificarne la conformità;
- le verifiche di processo per verificarne l'efficacia;
- le verifiche di prodotto, che sono indicate per frequenza nei rispettivi piani di controllo, per verificare la conformità a tutti i requisiti specificati e ai requisiti dimensionali, dove applicabile.

Durante il periodo di stage ho potuto affrontare diversi casi di controllo statistico del processo tra i quali uno in particolare verrà trattato in questo elaborato.

Un primo momento è stato dedicato all'ambientamento in azienda e alla conoscenza delle funzioni svolte in ogni reparto.

Successivamente, si è presentata la possibilità di poter operare nel Reparto di Rifinitura dove alle pelli viene applicato il colore richiesto dal cliente tramite la tecnica a spruzzo.

Per questo caso l'attenzione è stata posta sulla grammatura del colore cioè il peso del colore che viene applicato alle pelli.

Una seconda possibilità per applicare le mie conoscenze, è arrivata grazie al Reparto di Scelta dove le pelli vengono controllate prima della spedizione e, in particolare, sul controllo dello spessore delle pelli.

Ho potuto apprendere anche come l'azienda segue la taratura degli strumenti di misurazione come bilance, termometri e calibri quest'ultimi usati per misurare lo spessore delle pelli.

1.3 PRESENTAZIONE DELLA TESI

L'elaborato presenta lo studio del calcolo degli indici di capacità per lo spessore delle pelli.

Per questo caso di studio sono state prelevate 25 pelli da un unico cartellino e misurato lo spessore con l'utilizzo del calibro.

Il problema che si è presentato con questo caso ha reso necessario l'applicazione del calcolo degli indici di capacità per dati non – normali.

CAPITOLO 2

IL CASO

2. ANALISI DELLO SPESSORE DELLE PELLI

Lo spessore della pelle è un valore importante per il cliente che acquista le pelli e poi le deve lavorare per creare il prodotto finale.

Ogni tipologia di articolo, pelle per auto e arredamento, ha uno spessore differente in quanto per essere lavorato necessita di caratteristiche particolari.

Nel nostro caso l'articolo analizzato è una partita di pelli destinate alla carrozzeria e precisamente sedili per auto.

Le pelli per ottenere il giusto spessore nel lungo ciclo di lavorazione passano al Reparto Rasatura il quale ha il compito di uniformare la pelle ad uno spessore determinato.

Tale compito è affidato ad un operatore addetto alla macchina di rasatura, il quale imposta lo spessore desiderato.

La verifica dello spessore della pelle avviene nel Reparto di Scelta dove gli operatori attraverso l'uso del calibro misurano lo spessore delle pelli.

Per la nostra analisi sono state prelevate 25 pelli da un unico cartellino e, per ogni pelle sono stati rilevati 10 valori dislocati in 10 punti diversi della pelle secondo uno schema preciso imposto dal manuale della qualità.

I dati hanno un valore target di 1.20 mm con tolleranza $-0.10/+0.10$. I limiti di specifica dunque sono pari a $[1.10:1.30]$.

2.1 ANALISI ESPLORATIVA DEI DATI

Il dataset è composto da 25 campioni di ampiezza 10, la variabile rilevata è lo spessore delle pelli (variabile quantitativa continua).

I dati rappresentati nel grafico (Figura 2.1) mostrano le singole misure attraverso un istogramma di frequenza.

I limiti di specifica, corrispondono alle linee blu, mentre la linea rossa rappresenta il valore target dell'intervallo di specifica.

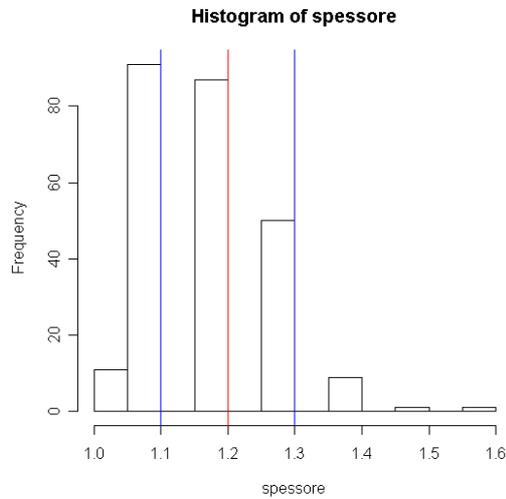


Figura 2.1: Istogramma di frequenza dello spessore delle pelli

Dal grafico si nota che il processo lavora molto vicino ai limiti di specifica e inoltre sono presenti dei valori che cadono fuori dai limiti di specifica.

La percentuale complessiva dei dati che cadono fuori dai limiti di specifica è pari all'8%.

Analizzando il boxplot della variabile di interesse si può notare che sono presenti dei valori anomali e che la mediana e il terzo quartile coincidono.

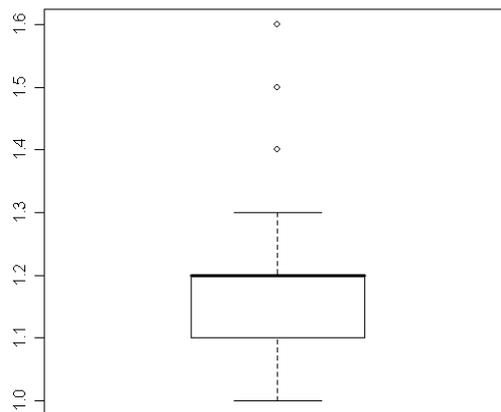


Figura 2.1.2: Boxplot dei dati

L'istogramma sulle osservazioni e il *normal probability plot* mettono in risalto che i dati non seguono la distribuzione normale.

Nell'istogramma la curva della densità non presenta il classico andamento a “campana” segno di normalità dei dati e il normal probability plot presenta un andamento che non approssima una retta con code pesanti che cadono fuori dai limiti -2 e 2.

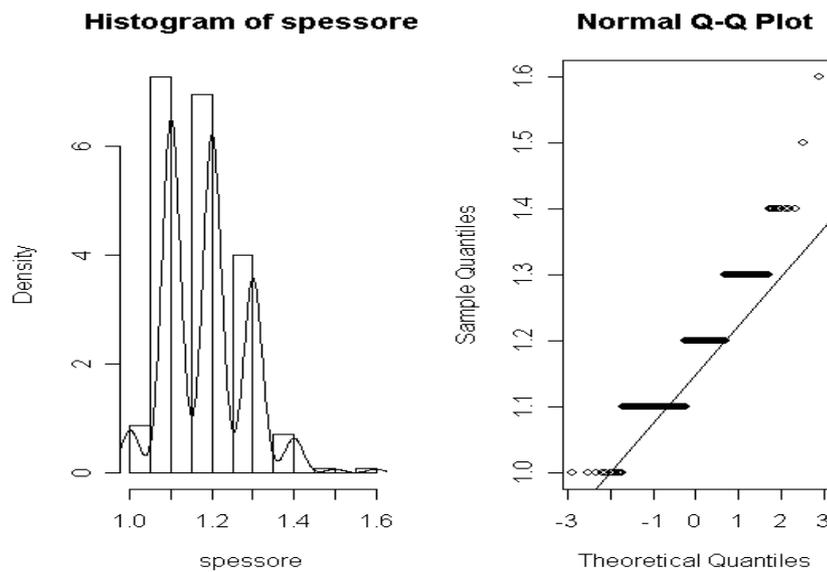


Figura 2.1.3: Istogramma e Normal Probability Plot dei dati

Per verificare la normalità dei dati con un test si è utilizzato il test di Shapiro-Wilk il quale porta a rifiutare l'ipotesi di normalità con un p-value pari circa a zero.

Essendo interessati a condurre un'analisi sulle medie per campione verifico l'eventuale normalità delle medie.

Dal boxplot emerge che la distribuzione delle medie presenta una coda lunga a destra e un'ampia variabilità.

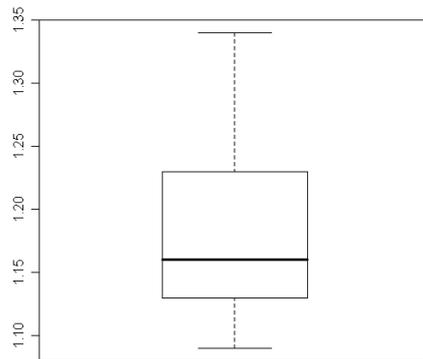


Figura 2.1.4: Boxplot delle medie per campione

L'istogramma e il normal probability plot mostrano un miglioramento dell'adattamento dei dati alla normale.

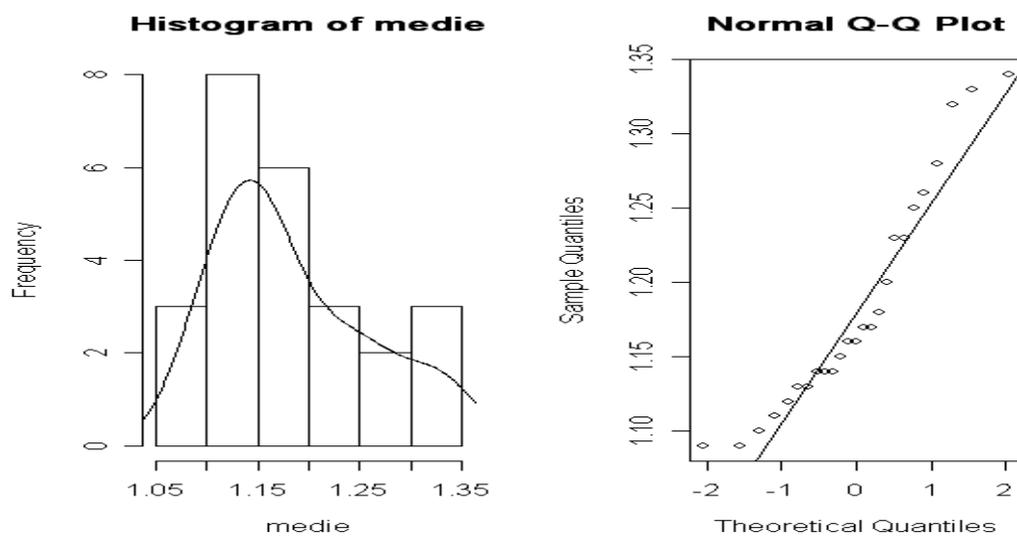


Figura 2.1.5: Istogramma e Normal Probability Plot delle medie per campione

Ma con il test di Shapiro-Wilk si giunge al rifiuto dell'ipotesi di normalità delle medie (p-value = 0.028).

Analizziamo ora l'autocorrelazione delle medie per campione.

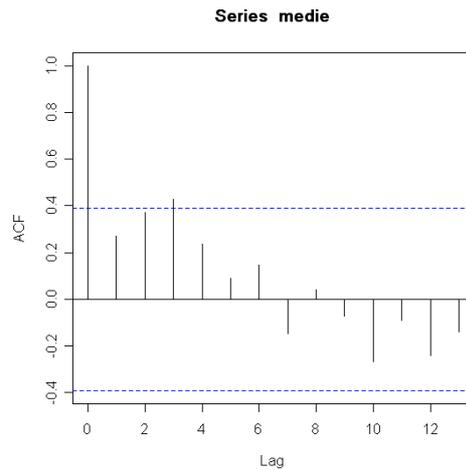


Figura 2.1.6: Grafio dell'autocorrelazione delle medie per campione

Il grafico mostra la presenza di un'autocorrelazione tra le medie e quindi risulta violata anche l'assunzione di indipendenza tra le osservazioni.

Il modello di serie storiche tipicamente utilizzato è il modello autoregressivo integrato a media mobile ARIMA(p,d,q). (Box e Jenkins,1976)

Volendo adattare un modello AR(1) alle osservazioni, dal test di Box Ljung si ottiene un p-value pari a 0.5404, che porta ad accettare la bontà del modello.

Un'ulteriore conferma arriva dalla Figure 2.1.7 e 2.1.8 che mostrano l'assenza di autocorrelazione e la normalità dei residui, rispettivamente.

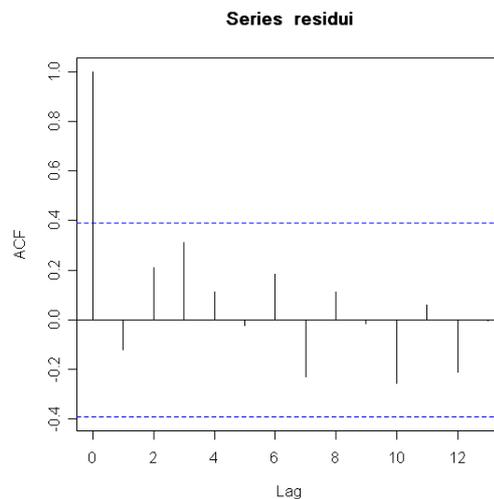


Figura 2.1.7: Grafico dell'autocorrelazione dei residui stimati con un modello AR(1)

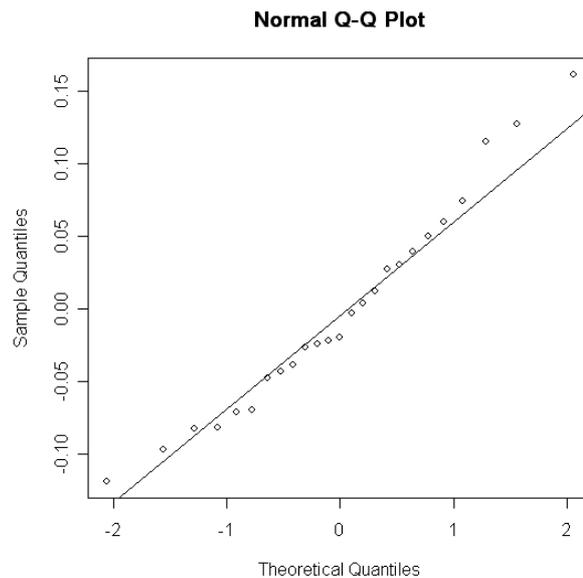


Figura 2.1.8: Normal Probability Plot dei residui di un AR(1)

Dopo questa prima analisi si può concludere che siamo in presenza di dati distribuiti non normalmente e autocorrelati, quindi si dovranno utilizzare carte di controllo opportune per questo caso come la carta con la statistica EWMA come linea centrale e la carta Shewhart per gli errori di previsione a un passo.

2.2 INTRODUZIONE ALLE CARTE DI CONTROLLO

Le ragioni di tale successo delle carte di controllo possono essere così riassunte:

1. Sono una comprovata tecnica per migliorare la produttività: un efficiente controllo della qualità riduce gli sprechi e la necessità di dover ripetere un lavoro
2. Sono efficaci per prevenire la produzione di pezzi difettosi
3. Evitano di dover apportare inutili aggiustamenti sul processo produttivo: una carta di controllo riesce a mettere in evidenza la presenza di variazioni anomale
4. Forniscono informazioni diagnostiche
5. Forniscono informazioni sulla capacità del processo e sulla sua stabilità nel tempo

La valutazione della qualità del processo si basa sul controllo delle sue caratteristiche e in particolare di alcune statistiche ad esse relative.

Ogni processo produttivo è soggetto ad una certa variabilità la quale può avere effetti diversi sulla variabilità del processo.

Si possono così suddividere le cause accidentali dalle cause identificabili o determinabili.

Un processo la cui variabilità sia provocata solo da **fattori accidentali** viene detto **sotto controllo**. Un processo che stia funzionando, invece, in presenza di **fattori identificabili** viene detto **fuori controllo**.

Una tipica carta di controllo riporta una linea centrale (LC) che rappresenta il valor medio della statistica di controllo e due linee orizzontali chiamate limite di controllo superiore (UCL) e limite di controllo inferiore (LCL).

Questi limiti di controllo vengono scelti in modo tale che, se il processo è sotto controllo, quasi tutti i valori campionari cadranno al loro interno e nessun intervento correttivo sul processo sarà necessario. Se invece un punto cade al di fuori di tali limiti, questo dovrà essere interpretato come un'evidenza del fatto che il processo è fuori controllo e quindi si renderanno necessarie azioni correttive o di indagine sul processo per individuare ed eliminare le cause che hanno portato all'insorgere di quei fattori specifici.

Le carte di controllo possono essere suddivise in due categorie, a seconda della caratteristica della variabile di studio.

Quindi possiamo utilizzare carte di controllo per variabili o carte di controllo per attributi.

Si utilizzano le **carte di controllo per variabili** quando la caratteristica da controllare è una variabile quantitativa.

Nel caso in cui la caratteristica da controllare sia di tipo qualitativo (si valuta la conformità dell'unità prodotta a seconda che possieda o meno certi attributi o a seconda del numero di difetti) vengono utilizzate le **carte di controllo per attributi**.

2.2.1 LA CARTA SHEWHART

Le carte di controllo dette **carte di Shewhart** fanno riferimento alla metodologia sviluppata da Walter A. Shewhart (Shewhart, 1931). Caratteristica principale delle carte di Shewhart è che esse fanno uso unicamente delle informazioni sul processo contenute

nell'ultimo istante di osservazione, ignorando tutte quelle precedenti risultando così insensibili a piccole modifiche nel livello del processo.

Le carte Shewhart per questo motivo vengono classificate tra le carte di controllo senza memoria.

Per poter applicare una carta Shewhart, le assunzioni di base che devono essere soddisfatte sono l'indipendenza e la normalità delle osservazioni.

Quando l'assunzione di indipendenza e di normalità è violata, come nel nostro caso, sono più opportune carte robuste alla non normalità delle singole misure e che considerino la correlazione tra le osservazioni passate.

2.2.2 LA CARTA EWMA

Le carte di controllo a medie mobili pesate esponenzialmente (EWMA) costituiscono una valida alternativa alle carte di Shewhart nel caso si sia interessati a scoprire salti di livello di entità limitata.

Le proprietà della carta EWMA possono essere così riassunte:

- Robusta alla non normalità delle singole misure
- Considera la correlazione tra osservazioni passate
- Evidenzia trend di deriva anche con poche osservazioni
- Ideale per carte per misure singole

La statistica di controllo consiste in una media mobile ponderata delle osservazioni passate, dove viene assegnato a ciascun valore un peso:

$$w_t = \sum_{j=0}^{t-1} \lambda(1-\lambda)^j x_{t-j} + (1-\lambda)^t w_0$$

dove $0 < \lambda \leq 1$ è una costante e il valore iniziale (richiesto con il primo campione per $i=1$) va fissato al valore di riferimento $w_0 = \mu_0$.

Se le osservazioni x_i sono determinazioni di variabili casuali indipendenti di varianza comune σ^2 , la varianza di w_t è pari a :

$$\sigma_w^2 = \sigma_0^2 \left(\frac{\lambda}{2-\lambda} \right) [1 - (1-\lambda)^{2t}]$$

ma, dal momento che, $1 - (1-\lambda)^{2t} \rightarrow 1$, la varianza asintotica della statistica EWMA diventa per $t \rightarrow \infty$, $\sigma_{wt}^2 = \sigma^2 (\lambda/2 - \lambda)$

I limiti di controllo della carta EWMA sono definiti come:

- Variabili per valori piccoli di t

$$LCL = \mu_0 - L \sigma_0 \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]}$$

$$LC = \mu_0$$

$$UCL = \mu_0 + L \sigma_0 \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2t}]}$$

- Si stabilizzano intorno ai valori asintotici, al crescere di t

$$LCL = \mu_0 - L \sigma_0 \sqrt{\frac{\lambda}{2-\lambda}}$$

$$LC = \mu_0$$

$$UCL = \mu_0 + L \sigma_0 \sqrt{\frac{\lambda}{2-\lambda}}$$

Se la media e la deviazione standard del processo sono incognite i limiti di controllo asintotici della carta EWMA diventano:

$$LCL = \bar{x} - L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LC = \bar{x}$$

$$UCL = \bar{x} + L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$\bar{x} = \frac{\sum x_i}{n}$$

$$\hat{\sigma} = \frac{MR(2)}{1.128}$$

$$\hat{\sigma} = s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Nel nostro caso si dispone di campioni di ampiezza maggiore di 1 quindi i limiti di controllo asintotici, con media e deviazione standard incognite, diventano:

$$LCL = \bar{\bar{x}} - L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LC = \bar{\bar{x}}$$

$$UCL = \bar{\bar{x}} + L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$\bar{\bar{x}} = \frac{\sum x_i}{n}$$

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

2.2.3 DISEGNO DELLA CARTA DI CONTROLLO EWMA

Le prestazioni di una carta EWMA sono valutate in termini di ARL (Average Run Length) ovvero il tempo di attesa medio per segnalare un allarme.

L'ARL(δ) con $\delta = \frac{\mu - \mu_0}{\sigma_0}$ può calcolarsi risolvendo in modo numerico un'equazione

integrale del tipo Fredholm (Crowder, 1989) o per mezzo dell'approccio Catene di Markov (Brook e Evans, 1971).

Inoltre, Crowder (1987), Lucas e Saccucci (1990), forniscono tavole dell'ARL della carta EWMA per diverse coppie di valori (λ , L).

Il disegno ottimale dello schema EWMA consiste pertanto nel trovare la combinazione ottima (λ^* , L^*) che produce, per un dato ARL in controllo, l'ARL fuori controllo più piccolo possibile, per un valore di δ specificato.

2.2.4 CARTA EWMA COME PREVISORE DEL LIVELLO MEDIO DEL PROCESSO

L'autocorrelazione è spesso un fenomeno naturale intrinseco di molti processi produttivi dove alcune variabili di processo variano lentamente rispetto al tasso al quale vengono misurate.

Nell'ambito dell'SPC (*Statistical Process Control*) l'autocorrelazione viene vista come un problema dal momento che porta ad interpretazioni fuorvianti dei segnali prodotti dalla carta Shewhart.

In letteratura sono presenti varie strategie per trattare l'autocorrelazione, una di queste suggerisce la rimozione dell'autocorrelazione dei dati e la costruzione di una carta Shewhart o EWMA o CUSUM per i residui. (*Alwan and Roberts 1988*)

Alwan e Roberts (1989), propongono di adattare un appropriato modello di serie storiche ai dati osservati, nel nostro caso è stato adattato un modello AR(1) che risulta efficiente se applicato ad una Shewhart o ad un'EWMA sui residui della serie.

Il modello di serie storiche tipicamente utilizzato è il modello autoregressivo integrato a media mobile ARIMA(p,d,q) *Box e Jenkins, 1976*.

Quando i dati sono correlati e possono essere modellati per mezzo di un ARIMA(p,d,q), la carta EWMA fornisce un buon previsore ad un passo.

La statistica di controllo w_t ora diventa \hat{X}_{t+1} ovvero la previsione ad un passo, fatta al periodo t per il periodo t+1

$$\hat{X}_{t+1}(t) = w_t = w_{t-1} + \lambda[X_t - \hat{X}_t(t-1)]$$

da cui si ricava

$$\hat{X}_{t+1}(t) = w_t = w_{t-1} + \lambda[X_t - \hat{X}_t(t-1)] = w_{t-1} + \lambda e_t$$

dove $e_t = X_t - \hat{X}_t(t-1) = X_t - w_{t-1}$ è l'errore di previsione ad un passo.

2.2.5 DISEGNO DI UN'EWMA PER DATI AUTOCORRELATI

Per considerare l'EWMA un buon previsore ad un passo, si pone il problema di come scegliere la costante λ . Il parametro λ può calcolarsi in modo da rendere minima la

somma dei quadrati degli errori di previsione ad un passo (ciò assicura che siano incorrelati) $\min_{\lambda} \sum_{t=1}^n e_t^2$

Un altro metodo che si può utilizzare è stato proposto da Cox (1961) che dimostra che con un AR(1) di parametro ϕ , l'EWMA fornisce un errore quadratico di previsione, ad un passo, minimo se $\lambda^{(1)} = 1 - \frac{1}{2} \left(\frac{1-\phi}{\phi} \right), \frac{1}{3} \leq \phi \leq 1$

Il parametro λ può anche scegliersi in modo da minimizzare la differenza al quadrato tra i pesi assegnati al passato, dalla procedura EWMA e quelli dati dal processo AR(1)

$$\lambda^{(2)} = 2 - \sqrt{\frac{1}{\phi}}$$

ϕ	0.25	0.33	0.4	0.5	0.6	0.7	0.8	0.9	0.95
$\lambda^{(1)}$		0	0.25	0.5	0.667	0.786	0.875	0.944	0.974
$\lambda^{(2)}$	0	0.259	0.419	0.586	0.709	0.805	0.882	0.946	0.974

Tabella 2.2.4.1: Valori per lambda

Si può osservare che:

- al crescere di ϕ cresce anche λ
- per $\phi \geq 0.6$, $\lambda^{(1)} \cong \lambda^{(2)}$
- per valori grandi di ϕ , una scelta ragionevole è $\lambda = \phi$

Dal momento che, i nostri dati non soddisfano l'ipotesi di normalità e di indipendenza, si considera una carta di controllo sugli errori di previsione ad un passo, tale carta è una carta per misure singole avente linea centrale pari a zero e limiti di controllo $\pm 3\sigma_p$

dove $\sigma_p = \sqrt{\frac{\sum_{t=1}^n e_t^2}{n}}$

Inoltre si propone di applicare una carta EWMA center line, dal momento che l'EWMA è un buon previsore ad un passo per il periodo t+1, avente:

- come linea centrale il valore della statistica EWMA $w_t = \hat{X}_{t+1}$

$$LCL_{t+1} = w_t - Z_{\frac{\alpha}{2}} \sigma_p$$

- come limiti di controllo

$$LC = w_t$$

$$UCL = w_t + Z_{\frac{\alpha}{2}} \sigma_p$$

- i valori osservati X_{t+1} riportati su tale carta

2.3 APPLICAZIONE DELLE CARTE DI CONTROLLO AI DATI

La costante lambda della carta EWMA standard da applicare alle medie per campione viene scelta applicando il criterio di Cox (1961) descritto in precedenza. Dalla Tabella 2.2.4.1 si ricava che, per un $\phi = 0.2994$ il valore di $\lambda^{(2)}$ è pari a 0.25.

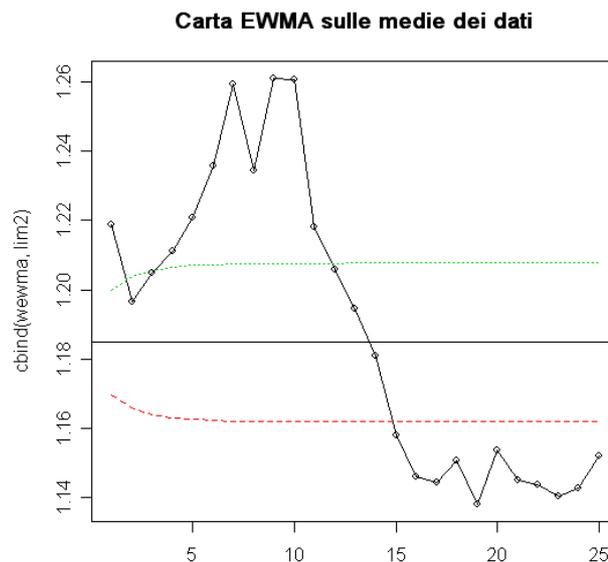


Figura 2.3.1: Carta Ewma sulle medie dei dati

La carta di controllo sulle medie segnala degli evidenti fuori controllo e una deriva decrescente che si stabilizza fuori dal limite inferiore di controllo.

Dal momento che, i residui ottenuti modellando i nostri dati, risultano distribuiti in modo normale è interessante costruire un'EWMA applicata ai residui di un modello AR(1).

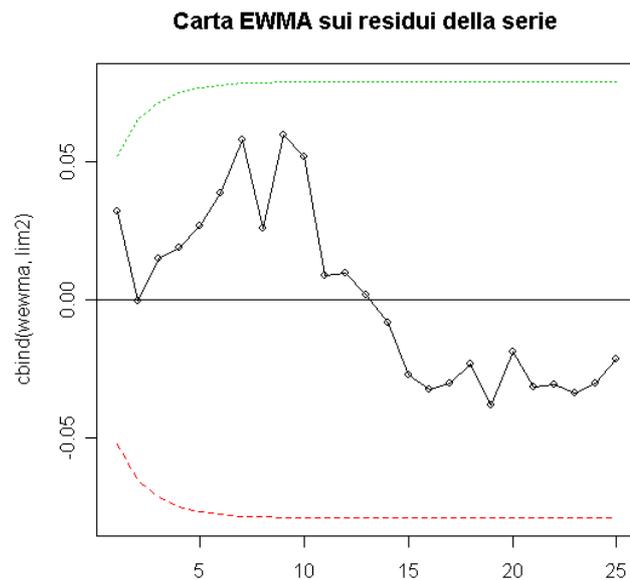


Figura 2.3.2: Carta EWMA sui residui della serie

La carta EWMA sui residui della serie mette in evidenza che si è in presenza di una qualche causa speciale in quanto i residui in un primo momento rimangono sopra la media e successivamente interviene una deriva negativa che porta i residui a stabilizzarsi sotto la media.

Applichiamo ora la carta Shewhart agli errori di previsione ottenuti attraverso la statistica EWMA. Per il disegno di questa carta si deve per prima cosa calcolare il valore di lambda cui corrisponde il più piccolo valore della somma dei residui al quadrato e la stima della varianza dell'errore basata su tale lambda (σ_p).

La carta Shewhart sugli errori di previsione che viene costruita è una carta sulle medie. Il valore di lambda ottimale risulta pari a 0.3 che corrisponde ad una somma dei residui al quadrato pari a 0.119 e la stima di σ_p che si ottiene è 0.069.

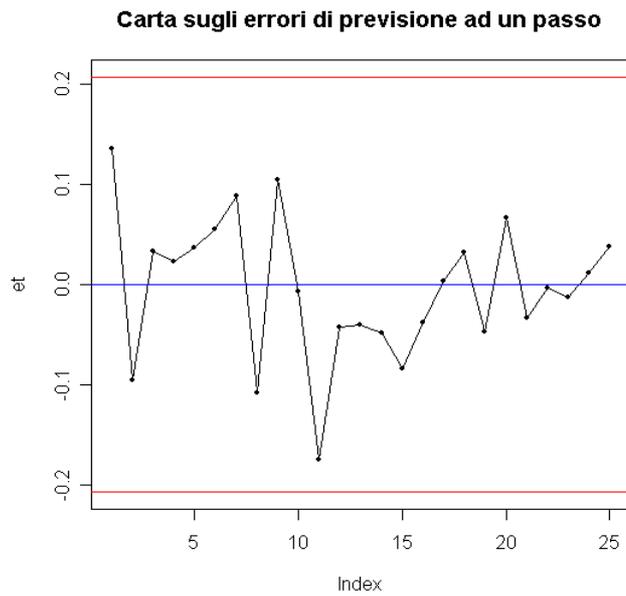


Figura 2.3.2: Carta Shewhart sugli errori di previsione a un passo

Il grafico rimane in controllo.

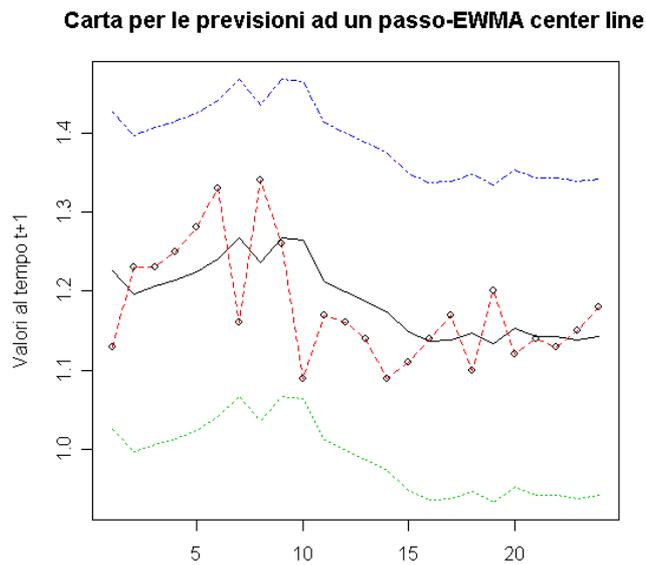


Figura 2.3.3: Carta EWMA center line

Dal grafico si nota che il processo ad un certo punto subisce una modifica e l'andamento dei dati si stabilizza intorno alla media dopo l'intervento dell'operatore.

CAPITOLO 3

CARTE DI CONTROLLO PER DATI NON - NORMALI

3.1 CARTE DI CONTROLLO BASATE SUI RANGHI

Le carte di controllo usuali sono usate per individuare se un processo rimane in uno stato di controllo statistico. Questo per valutare se la distribuzione del processo o dei suoi parametri come la media, rimane nei livelli specificati.

E' noto che le carte di controllo come la Shewhart e la CUSUM sono molto sensibili alla violazione dell'ipotesi di normalità e quindi poco adatte a rappresentare quei processi che presentano una distribuzione non-normale.

Le carte di controllo non parametriche sono una pratica alternativa alle carte di controllo standard quando la distribuzione dei dati è non-normale.

Hackl e Ledolter (1991) introducono le carte EWMA basate sui ranghi, dove il rango dell'osservazione si basa sulle più recenti g osservazioni.

Questa carta è efficace rispetto a trend lineari del processo, ma è piuttosto poco robusta in presenza di salti nella media.

3.2 RANGHI STANDARDIZZATI

Si consideri $X_t, t=1,2,\dots$, una sequenza di variabili indipendenti da una distribuzione di probabilità continua $F(x)$. Il problema di tenere sotto controllo il processo deriva dal dover identificare veloci cambiamenti in $F(x)$. Il processo è considerato sotto controllo se $F(x)=F_0(x)$.

I ranghi standardizzati R_t delle osservazioni X_t della distribuzione in controllo vengono così definiti:

$$R_t = 2 \left[F_0(X_t) - \frac{1}{2} \right] \quad (1)$$

Le proprietà di R_t sono facilmente derivabili e si può mostrare che R_t si distribuisce in modo uniforme tra -1 e 1, con media zero e varianza $\frac{1}{3}$, e i ranghi tra loro sono indipendenti.

Se la distribuzione in controllo non è specificata si può usare un campione di dati di ampiezza $g-1$, $\{Y_1, Y_2, \dots\}$, estratto dal processo in controllo, e si definiscono i ranghi standardizzati di X_t come:

$$R_t = \frac{2}{g} \left(R_t^* - \frac{g+1}{2} \right) \quad (2)$$

dove R_t^* è il rango di X_t rispetto a $\{Y_1, Y_2, \dots\}$ e

$$R_t^* = 1 + \sum_{i=1}^{g-1} I(X_t > Y_i)$$

con $I(X_t > Y_i) = 1$ se $X_t > Y_i$, e zero altrimenti.

I ranghi nella (2) seguono una distribuzione uniforme discreta nei punti di g con media zero e varianza $(g^2 - 1)/3g^2$.

Usando la funzione indicatrice possiamo vedere che la correlazione $\text{Corr}\{R_t, R_{t+k}\} = (g+1)^{-1}$ per tutti i valori di k .

Se g inizia ad aumentare la correlazione tende a zero e si torna ai ranghi standardizzati definiti nell'equazione (1).

3.3 EXPONENTIALLY WEIGHTED MOVING AVERAGES

Per controllare la media del processo viene proposta una carta EWMA basata sui ranghi.

La statistica di controllo è definita come:

$$T_t = (1 - \lambda)T_{t-1} + \lambda R_t \quad (3)$$

per ogni periodo $t=1,2,\dots$

Il valore di partenza T_0 è di solito posto uguale a zero e λ , con $0 < \lambda \leq 1$, è il parametro di liscio.

Per il parametro λ è di solito raccomandato un valore compreso tra 0.1 e 0.3, per $\lambda=0.3$, per esempio, l'88% della statistica è determinato dal peso delle ultime 6 osservazioni.

Il processo è considerato fuori controllo al tempo t , se $T_t > h$, dove h è un conveniente limite di controllo opportunamente scelto.

Nell'usare la carta EWMA basata sui ranghi si deve tenere presente che, tale carta lascia prevedere che si possa perdere qualcosa in termini di efficienza se la variabile di controllo si distribuisce come una normale, ma risulta robusta nel caso in cui l'ipotesi di normalità è violata.

Tale carta dipende da due parametri: il parametro di lisciamento λ e il limite di controllo h .

La prestazione del processo in controllo è valutata in termini di ARL, cioè il numero medio di osservazioni che sono necessarie per individuare un fuori controllo.

I parametri λ e h sono scelti in modo da avere un ARL il più grande possibile se il processo è in controllo e il più piccolo possibile nel caso di fuori controllo.

Prendendo in considerazione un valore per h compreso tra 0.50 e 0.60 e $\lambda = 0.3$ fornisce un ARL tra 100 e 430 se la distribuzione di riferimento è nota.

In caso contrario, se la distribuzione è ignota, l'ARL è posto tra 200 e 1050 con un campione di riferimento di ampiezza 50.

La carta EWMA basata sui ranghi riduce la sua sensibilità in presenza di outliers, ciò comporta che la carta non risulta sensibile a grandi cambiamenti nella media.

3.4 APPLICAZIONE DELLA CARTA EWMA BASATA SUI RANGHI

I dati a cui ci riferiamo sono gli spessori raccolti per 25 pelli e per le quali abbiamo 10 misurazioni per pelle.

Non avendo alcun campione di riferimento, vengono prese le misurazione delle prime 10 pelli come campione di confronto per il calcolo dei ranghi standardizzati sulla media.

I parametri per il disegno della carta EWMA sui ranghi sono stati scelti seguendo la procedura proposta da Hackl e Ledolter, i quali dimostrano che per un $\lambda = 0.3$, un $h = 0.5$ e un numero di osservazioni in controllo inferiore a 50, il numero minimo di osservazioni per osservare un fuori controllo è 3.

3.4.1 ANALISI PRELIMINARE DEI DATI

Volendo analizzare i dati divisi nei due campioni, costruiamo i relativi boxplot dai quali si nota che, per il primo campione di riferimento la media è 1.25 mentre per il secondo campione la media è circa 1.14, quindi una differenza in media pari a 0.11.

I limiti di specifica, come già detto in precedenza, sono: [1.10;1.30].

Per quanto riguarda la variabilità i due grafici sono molto simili.

Boxplot dei due campioni

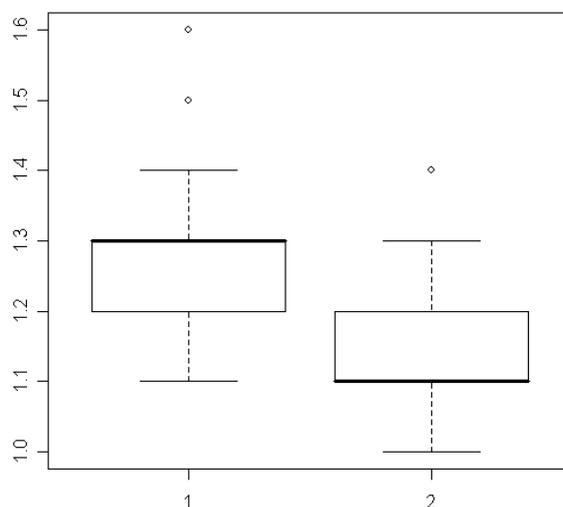


Figura 3.4.1.1: Boxplot relativi ai dati divisi per campione

La carta basata sui ranghi come ipotesi prevede che la distribuzione dei dati non sia normale, quindi costruiamo il normal probability plot per ciascun campione.

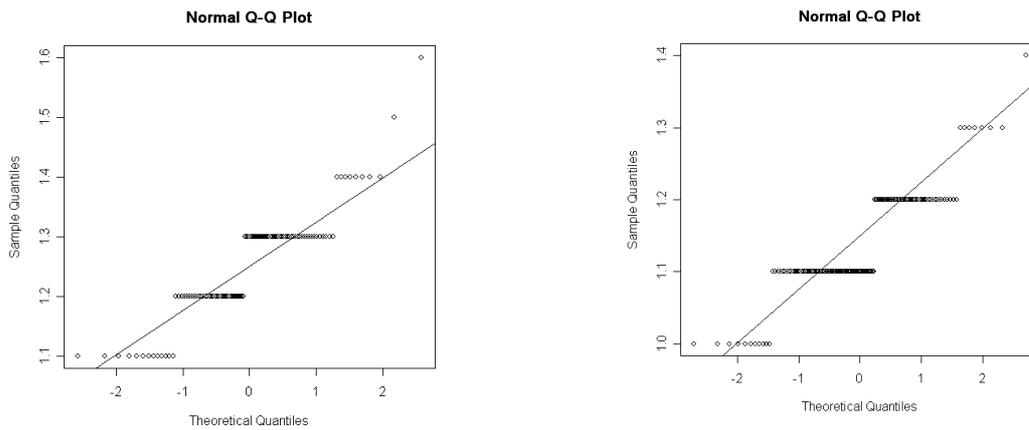


Figura 3.4.1.2: Normal Probability Plot dei dati divisi per campione

Dai grafici si può vedere come i dati non provengano da una distribuzione normale in quanto entrambi non si approssimano ad una retta.

3.4.2 CARTA EWMA BASATA SUI RANGHI

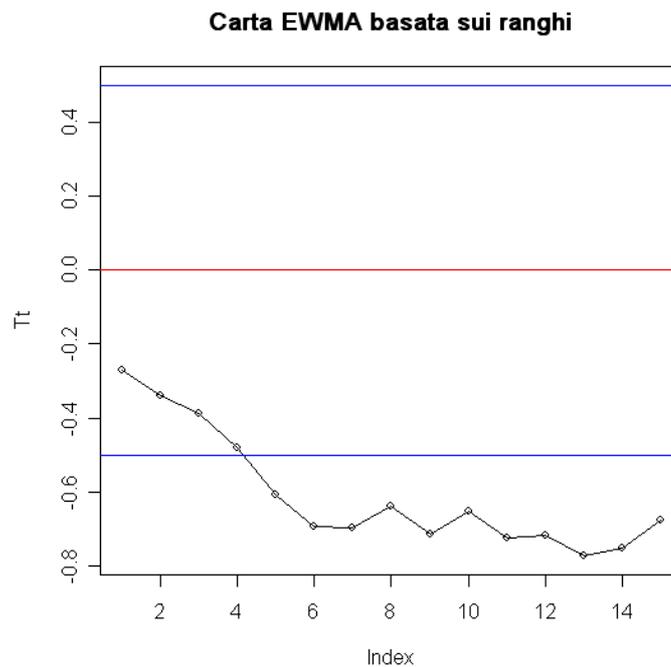


Figura 3.4.2.1: Carta EWMA basata sui ranghi per $\lambda=0.3$ e $h=0.5$

Dal grafico si può notare come, i ranghi calcolati sulle medie, riescano a descrivere l'andamento dei dati.

Il processo dopo la quarta osservazione segnala un fuori controllo e poi si stabilizza fuori dal limite inferiore di controllo, si suppone che questo sia dovuto al fatto che il processo tenda a lavorare molto vicino ai limiti di specifica, e l'operatore tarda ad intervenire.

CAPITOLO 4

STUDIO DELLA CAPACITA' DEL PROCESSO

4. ANALISI DELLA CAPACITA' DEL PROCESSO

L'analisi della capacità del processo può essere definita come l'analisi della variabilità naturale del processo rispetto alle specifiche del prodotto.

Le tecniche statistiche sono utili durante tutto il ciclo di produzione, a partire dalle attività di sviluppo preliminari alla fabbricazione vera e propria, per continuare con l'analisi e la valutazione della variabilità del processo, in relazione ai livelli nominali di specifica, fino alle operazioni dedicate alla eliminazione o almeno riduzione di detta variabilità.

La capacità del processo viene in generale riferita alla uniformità di comportamento del processo.

La variabilità misura l'uniformità delle caratteristiche del prodotto, questa variabilità può essere interpretata in due modi:

- variabilità istantanea o naturale relativa ad un determinato periodo
- variabilità nel tempo misurata in diversi istanti temporali

La performance corrente è misurata dalla variabilità del processo (espressa in deviazioni standard) e dalle deviazioni della media del processo dal livello nominale (quando questo non è specificato dal centro dell'intervallo definito dai limiti di specifica).

4.1 GLI INDICI DI CAPACITA'

Gli indici di capacità sono considerati uno degli strumenti più importanti per misurare la qualità per un continuo miglioramento della qualità e della produttività.

Il primo indice che andiamo ad analizzare è l'indice C_p che misura la capacità potenziale del processo a breve termine.

L'indice C_p pone a confronto l'ampiezza della specifica con il 6σ della variabilità.

Si può definire il C_p come:
$$C_p = \frac{USL - LSL}{6\sigma_{ST}}$$

dove USL e LSL sono rispettivamente i limiti superiore ed inferiore di specifica, σ_{ST} è la deviazione standard “short term” che viene stimata attraverso $\hat{\sigma}$. La stima di $\hat{\sigma}$ viene effettuata utilizzando il range medio nel caso di campioni di ampiezza maggiore di 1

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

mentre si utilizza l’escursione mobile nel caso di osservazioni singole.

$$\hat{\sigma} = \frac{\overline{MR(2)}}{d_2}$$

dove d_2 è una costante in funzione dell’ampiezza campionaria.

Il valore di C_p più comunemente usato come limite inferiore per determinare se un processo sia capace è $C_p = 1.33$. L’unico limite che si riscontra nell’indice C_p è che non tiene conto di dove si posiziona la media rispetto ai limiti di specifica, per questo motivo è stato introdotto l’indice C_{pk} che tiene conto della centratura del processo.

L’indice C_{pk} misura la capacità effettiva del processo e viene calcolato come segue:

$$C_{pk} = \frac{\min(USL - \mu, \mu - LSL)}{3\sigma_{ST}}$$

dove μ è la stima della media dei valori osservati.

Nel caso di singole osservazioni $\mu = \bar{x}$ che è la media di tutte le osservazioni, mentre nel caso di campioni di ampiezza maggiore di 1 $\mu = \bar{\bar{x}}$ che rappresenta la media delle medie campionarie.

I valori che assume il C_{pk} possono essere così interpretati:

- $C_{pk} < 0$ la media giace fuori dai limiti di specifica
- $C_{pk} = 0$ processo centrato su uno dei limiti di specifica
- $0 < C_{pk} < 1$ processo poco capace
- $1 < C_{pk} < 1.66$ processo marginalmente capace
- $C_{pk} > 1.66$ processo capace

In generale, se $C_p = C_{pk}$, il processo è centrato rispetto all’intervallo di specifica, mentre quando $C_{pk} < C_p$ il processo non è centrato.

Nell’interpretare questi indici è importante tenere presente che l’ipotesi di normalità per la distribuzione della variabile in studio sia soddisfatta.

L'indice di capacità C_{pk} è stato inizialmente sviluppato per il fatto che il C_p non trattava sufficientemente bene il caso di processi con media μ non centrata all'interno dei limiti di specifica. Tuttavia, il solo C_{pk} risulta anch'esso inadeguato come misura di centratura del processo.

Per ogni valore prefissato di μ nell'intervallo tra LSL e USL, l'indice C_{pk} dipende in maniera inversamente proporzionale da σ e diventa elevato per valori di σ prossimi allo zero. Questa caratteristica può rendere il C_{pk} inadatto come misura di centratura in quanto un valore elevato di C_{pk} non fornisce informazioni sulla posizione della media del processo nell'intervallo tra LSL e USL.

Un modo per superare questa difficoltà è usare un indice di capacità che sia un indicatore migliore della centratura, uno di questi è il C_{pmk} che viene definito come:

$$C_{pmk} = \frac{\min(USL - \bar{X}, \bar{X} - LSL)}{3\tau}$$

dove τ è la radice quadrata della deviazione attesa rispetto al valore di riferimento $T = (USL + LSL)/2$:

$$\tau^2 = \sigma^2 + (\mu - T)^2$$

Boyles (1991) osserva che sia C_{pk} sia C_{pmk} coincidono con C_p quando $\mu = T$ e decrescono man mano che μ si allontana da T . Un valore negativo di C_{pmk} significa che il processo lavora fuori dai limiti di specifica.

Questo indice possiede una elevata sensibilità alle situazioni di allontanamento della media del processo dal valore obiettivo T desiderato.

L'indice di capacità C_{pm} considera la distanza tra la media e il valore obiettivo ma non fornisce indicazioni se il livello medio del processo è interno ai limiti di specifica.

L'indice C_{pm} può essere scritto come

$$C_{pm} = \frac{USL - LSL}{6\tau}$$

4.2 INDICI DI CAPACITA' PER DATI NON- NORMALI

Gli indici di capacità fin ora descritti forniscono dei risultati accettabili se l'ipotesi di normalità dei dati è soddisfatta.

Se questi indici vengono usati per dati la cui distribuzione non è normale, i valori degli indici risultano incorretti e potrebbero distorcere l'attuale qualità del prodotto.

Choobineh e Ballare (1987) hanno introdotto per la prima volta il metodo della varianza pesata (weighted variance) per costruire le carte di controllo quando la distribuzione non è normale.

Il metodo della varianza pesata è stato utilizzato per modificare gli indici di capacità e poterli applicare così a quei dati che presentano una distribuzione non normale.

Gli unici indici ai quali si può applicare il metodo della varianza pesata sono: \hat{C}_p , \hat{C}_{pk} , \hat{C}_{pm} , \hat{C}_{pmk} .

L'idea che sta alla base del metodo di varianza pesata è quello di dividere la distribuzione asimmetrica in due distribuzioni normali con stessa media ma deviazione standard differente.

Per una popolazione con media μ e deviazione standard σ , ci sono n_1 osservazioni su n osservazioni totali che sono inferiori o uguali a μ e n_2 osservazioni su n osservazioni totali che sono maggiori di μ .

Le due nuove distribuzioni possono essere costruite usando n_1 e n_2 osservazioni rispettivamente e di conseguenza, le due nuove distribuzioni avranno la stessa media ma deviazioni standard differenti σ_1 e σ_2 .

Dal momento che, μ , σ_1 e σ_2 sono di solito sconosciuti è richiesta la stima dei parametri.

Per la stima di μ si ha $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$ mentre σ_1 e σ_2 possono essere stimati da S_1^2 e S_2^2 rispettivamente; dove S_1 è la stima della deviazione standard per le n_1 osservazioni che sono inferiori o uguali al valore di \bar{X} (interpretazione analoga per S_2).

$$\text{Quindi avremo: } S_1^2 = \frac{2 \sum_{i=1}^{n_1} (X_i - \bar{X})^2}{2n_1 - 1} \quad (4)$$

$$S_2^2 = \frac{2 \sum_{i=1}^{n_2} (X_i - \bar{X})^2}{2n_2 - 1} \quad (5)$$

La stima della deviazione standard dal valore target può essere espressa considerando lo stimatore di Taguchi

$$S_{T1}^2 = \frac{2 \sum_{i=1}^{n_1} (X_i - T)^2}{2n_1} \quad (6)$$

$$S_{T2}^2 = \frac{2 \sum_{i=1}^{n_2} (X_i - T)^2}{2n_2} \quad (7)$$

dove T è il valore target.

Gli indici di capacità vengono così modificati:

$$\hat{C}_p(WV) = \frac{USL - LSL}{3(S_1 + S_2)}$$

$$\hat{C}_{pk}(WV) = \min \left[\frac{USL - \bar{X}}{3S_2}, \frac{\bar{X} - LSL}{3S_1} \right]$$

dove S_1 e S_2 sono definiti nelle equazioni (4) e (5).

Gli indici $\hat{C}_{pm}(WV)$ e $\hat{C}_{pmk}(WV)$ possono essere definiti da:

$$\hat{C}_{pm}(WV) = \min \left[\frac{USL - T}{3S_{T2}}, \frac{T - LSL}{3S_{T1}} \right]$$

$$\hat{C}_{pmk}(WV) = \min \left[\frac{USL - \bar{X}}{3S_{T2}}, \frac{\bar{X} - LSL}{3S_{T1}} \right]$$

dove S_{T1} e S_{T2} sono definiti nelle equazioni (6) e (7).

Il vantaggio di usare il metodo della varianza pesata per modificare gli indici di capacità basati sulla normalità del processo è che gli indici basati sul metodo della varianza pesata riflettono il valore della distorsione della distribuzione.

4.3 INDICI DI CAPACITA' SUI DATI OSSERVATI

Calcoliamo ora gli indici di capacità per i dati osservati basati sulla varianza pesata per ovviare il problema della non-normalità dei dati.

Attraverso la funzione creata in R per il calcolo della varianza pesata, si sono ottenute le seguenti stime:

$$S_1 = 0.9677$$

$$S_2 = 0.8028$$

$$S_{T1} = 1.1189$$

$$S_{T2} = 0.6165$$

mentre per la media la stima che si è ottenuta è pari a 1.1848.

I valori che risultano per il C_p (0.0377) e il C_{pk} (0.0292) si avvicinano molto a zero, ciò significa che il processo lavora molto vicino ai limiti di specifica e non è molto capace.

Attraverso un grafico mostriamo come si distribuiscono i dati all'interno dei limiti di specifica e quanto la media dista dal valore target.

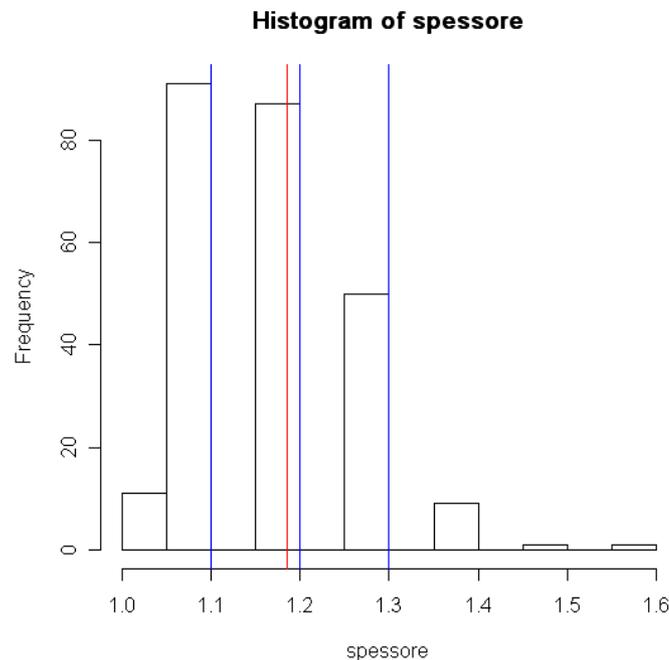


Figura 4.3.1: Istogramma dei dati con evidenza dei limiti di specifica

Nell'istogramma le linee verticali in blu indicano i limiti di specifica (1.10,1.30) e il valore target (1.20), mentre la linea verticale rossa indica la media calcolata sui dati osservati.

Si può notare come il processo lavori anche fuori dai limiti di specifica e, nonostante la media del processo si avvicini molto al valore target i valori degli indici Cpm (0.0298) e Cpmk (0.0253) sono inferiori a 1.

Si può concludere affermando che la qualità del prodotto non è sufficiente vista la grande variabilità dei dati anche fuori dai limiti di specifica.

CAPITOLO 5

CONCLUSIONI

5. CONCLUSIONI

Lo scopo di questa relazione era quello di analizzare lo spessore delle pelli con carte di controllo appropriate.

Da una prima analisi esplorativa dei dati si è notato che questi si distribuivano in modo non-normale e con valori che si concentravano nei limiti superiore e inferiore del processo.

Dal momento che, si era interessati ad un'analisi sulle medie per campione si è andati a valutare la possibile normalità delle medie con il risultato che nemmeno le medie si distribuivano come una normale.

Si è inoltre constatato che le medie presentavano una modesta autocorrelazione e si è deciso di applicare un modello autoregressivo di tipo AR(1).

Le carte di controllo che sono state utilizzate sono state scelte sulla base della non-normalità dei dati e per la loro autocorrelazione.

La prima carta che siamo andati a costruire è stata la carta Shewhart sugli errori di previsione ad un passo che, nel complesso risulta soddisfacente.

Successivamente si è costruita la carta EWMA center line dalla quale si è notato che i dati ad un certo istante subiscono una modifica per poi stabilizzarsi intorno alla media.

Una valida alternativa alle carte di controllo standard quando la distribuzione dei dati è non-normale sono le carte di controllo non parametriche in particolare, nel nostro caso, è stata adottata la carta EWMA basata sui ranghi applicata alle medie per campione.

Infine, sono stati calcolati gli indici di capacità del processo per dati non-normali dai quali si è arrivati alla conclusione che, il processo lavora vicino e fuori dai limiti di specifica con la conseguenza che il prodotto potrebbe risultare di qualità inferiore e il cliente insoddisfatto della merce acquistata.

Sarebbe quindi utile poter controllare lo spessore delle pelli con frequenza più elevata a scapito ovviamente di rallentare la fase di misurazione e di imballaggio della merce.

APPENDICE

Funzioni create per l'applicazione delle carte di controllo basate sui ranghi e per il calcolo degli indici di capacità per dati non-normali.

- Funzione per il calcolo dei ranghi sulla media per campioni di ampiezza > 1

```
rstar<-function(dati,l)
{
x<-wcamp(dati,l)[,1]
n<-length(x)
g<-x[1:((n-1)/3)]
d<-length(g)
f<-x[((n-1)/3)+1:n]
w<-c(rep(0,d))
rstar<-c(rep(0,length(f)))
r<-c(rep(0,length(f)))
for(t in 1:length(f))
{
  for (i in 1:d)
  {
    if (f[t]>g[i])
      w[i]<-1+sum(w[i-1])
    else
      w[i]<-0+sum(w[i-1])
    rstar[t]<-1+sum(w[i])
    i<-i+1
    r[t]<-(2/d)*(rstar[t]-((d+1)/2))
  }
t<-t+1
}
cbind(rstar,r)
}
```

- Funzione per il calcolo della statistica EWMA basata sui ranghi

```
EWMAranghi<-function(dati,lam,n)
{
r<-rstar(dati,n)[,2]
v<-c(rep(0,length(r)))
for (t in 1:length(r))
{
  if (t==1)
    v[t]<-lam*r[t]
  else
    v[t]<-((1-lam)*v[t-1])+(lam*r[t])
  t<-t+1
}
v
}
```

- Funzione per il calcolo della varianza pesata

```
vaWV<-function(dati,xmedio,vtarget)
{
n<-length(dati)
x2<-dati[dati>xmedio]
x1<-dati[dati<=xmedio]
n1<-length(x1)
n2<-length(x2)
S1<-sqrt(2*(sum(x1-xmedio)^2)/(2*n1-1))
cat("S1= ",round(S1,digits=4), "\n")
S2<-sqrt(2*(sum(x2-xmedio)^2)/(2*n2-1))
cat("S2= ", round(S2,digits=4), "\n")
ST1<-sqrt(2*(sum(x1-vtarget)^2)/(2*n1))
cat("ST1= ", round(ST1,digits=4), "\n")
ST2<-sqrt(2*(sum(x2-vtarget)^2)/(2*n2))
cat("ST2= ",round(ST2,digits=4), "\n")
}
```

- Funzione per il calcolo degli indici di capacità per dati non-normali

```
CapWV<-function(xmedio,S1,S2,ST1,ST2,vtarget)
{
LSL<-vtarget-0.1
USL<-vtarget+0.1
Cp<-(USL-LSL)/(3*(S1+S2))
cat("Cp= ",round(Cp,digits=4), "\n")
zlsl<-(xmedio-LSL)/(3*S1)
zusl<-(USL-xmedio)/(3*S2)
Cpk<-min(zlsl,zusl)
cat("Cpk= ",round(Cpk, digits=4), "\n")
vlsl<-(xmedio-LSL)/(3*ST1)
vusl<-(USL-xmedio)/(3*ST2)
Cpmk<-min(vlsl,vusl)
cat("Cpmk= ", round(Cpmk,digits=4),"\n")
wlsl<-(vtarget-LSL)/(3*ST1)
wusl<-(USL-vtarget)/(3*ST2)
Cpm<-min(wlsl,wusl)
cat("Cpm= ",round(Cpm,digits=4), "\n")
}
```

BIBLIOGRAFIA

- Hackl P., Ledolter J. (1991), “A control chart based on ranks” , *Journal of Quality Technology* , vol. 23
- Chakraborti S., Van Der Laan P., Bakir S. T., (2001), “ Nonparametric control charts: an overview and some results”, *Journal of Quality Technology*, vol. 33
- Wu H. H., Swain J., Farrington P., Messimer S. (1999), “A weighted variance capability index for general non-normal processes”, *Qualità and Reliability Engineering International*, vol.15, pp. 397-402
- Montgomery D. (2000), “Controllo statistico della qualità”, *McGraw-Hill*
- Shewhart, WA (1931), “Economic control of quality of manufactured product”, *D. Van Nostrand Company, Inc.*