



UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE
CORSO DI LAUREA IN
STATISTICA E TECNOLOGIE INFORMATICHE

TESI DI LAUREA TRIENNALE

**CARTE DI CONTROLLO SHEWHART E
CUSUM PER LE COMPONENTI DI VARIANZA**

Relatore: Ch.ma prof.ssa Giovanna Capizzi

Laureando: Massimiliano Gasperin

ANNO ACCADEMICO 2005/2006

INDICE

	Pag.
• INTRODUZIONE	1
• CAPITOLO 1: Generalità sulle componenti di varianza	3
○ 1.1. Generalità	3
○ 1.2. Esempio Reale	4
○ 1.3. Il modello	5
○ 1.4. Le sequenze di controllo	5
• CAPITOLO 2: Disegno delle componenti di varianza	9
○ 2.1. Carte di controllo Shewhart	9
○ 2.2. Disegno della carta Shewhart per il controllo di $\hat{\sigma}^2$	10
○ 2.3. Disegno della carta Shewhart per il controllo di $\hat{\sigma}_b^2$	10
○ 2.4. Disegno della carta Shewhart per il controllo di μ	12
○ 2.5. Carte di controllo CUSUM	13
○ 2.6. Disegno della carta CUSUM per il controllo di $\hat{\sigma}^2$	14
○ 2.7. Disegno della carta CUSUM per il controllo di $\hat{\sigma}_b^2$	14
○ 2.8. Disegno della carta CUSUM per il controllo di μ	15
• CAPITOLO 3: Costruzione delle carte di controllo	17
○ 3.1. Costruzione per le carte Shewhart	17
▪ 3.1.1. Componente $\hat{\sigma}^2$	17
▪ 3.1.2. Componente $\hat{\sigma}_b^2$	18
▪ 3.1.3. Componente μ	18
○ 3.2. Costruzione per le carte CUSUM	19
▪ 3.2.1. Componente $\hat{\sigma}^2$	19
▪ 3.2.2. Componente $\hat{\sigma}_b^2$	20
▪ 3.2.3. Componente μ	21

• CAPITOLO 4: Confronti tra le carte di controllo	23
○ 4.1. Confronto per la componente σ	23
○ 4.2. Confronto per la componente σ_b	25
○ 4.3. Confronto per la componente μ	28
• CAPITOLO 5: Performance delle carte di controllo	31
○ 5.1. Performance dei sistemi di monitoraggio	31
○ 5.2. Performance delle carte di controllo	32
• CAPITOLO 6: Carte tradizionali	39
○ 6.1. Carte Shewhart tradizionali	39
▪ 6.1.1. Carta basata su S^2	39
▪ 6.1.2. Carta per le medie	40
○ 6.2. Applicazione	41
▪ 6.2.1. Valutazioni	42
○ 6.3. Verifica della presenza di ulteriori fonti di variabilità	45
○ 6.4. Confronti tra le carte tramite i profili ARL	48
• CONCLUSIONI	53
• APPENDICE	55
• BIBLIOGRAFIA	61
• RINGRAZIAMENTI	63

INTRODUZIONE

Il controllo della varianza del processo è importante per assicurare la qualità consistente dei prodotti. In molti casi la variabilità globale dipende da diverse cause assegnabili. Quindi piuttosto che considerare un'unica fonte di varianza che determini quella totale, si possono considerare diverse componenti e, a partire da esse, costruire differenti carte di controllo per il loro monitoraggio. Questo articolo discute le tecniche Shewhart (Chang e Gan, 2004) e CUSUM (Yaschin, 1994) per il controllo basato su diverse componenti di variabilità, mettendo in evidenza la differenza con le carte di controllo usate assumendo l'esistenza di una sola fonte di varianza per il processo. Il primo capitolo affronterà un discorso introduttivo più ampio, prendendo in esame un caso reale, con la costruzione delle sequenze di controllo. Il secondo capitolo mostrerà il disegno delle carte di controllo, e il terzo capitolo la costruzione di esse tramite un esempio pratico. Successivamente il quarto capitolo confronterà le carte di controllo Shewhart e CUSUM con la decomposizione della varianza per valutare la loro efficienza. Il quinto capitolo confronterà le due tecniche tramite l'analisi della performance basata sull'ARL (average run length), mentre l'ultimo si occuperà di confrontare le carte Shewhart tradizionali e quelle discusse da Chang e Gan mettendo in risalto le differenze riscontrate. Infine seguiranno le conclusioni e un'appendice contenente i principali programmi Java per il calcolo delle ARL.

CAPITOLO 1

Generalità sulle componenti di varianza

1.1. Generalità

Nello studio del controllo della qualità, si è soliti fare riferimento a 2 principali carte per sorvegliare il processo: la carta Shewhart e la carta CUSUM. I parametri solitamente tenuti sotto controllo sono la media e la varianza delle osservazioni rilevate. Le carte tradizionali Shewhart, denominate carte senza memoria, controllano la media del processo mediante la stima della media sui campioni rilevati, e la variabilità con carte basate sulla deviazione standard o i range campionari (dipende dalla grandezza dei campioni). Diversi studi di queste carte si possono trovare su molti libri, tra i quali citiamo Duncan (1986), Grant e Leavenworth (1988), e Montgomery (2001). Le tecniche CUSUM si basano essenzialmente sul controllo di statistiche che tengano conto dei parametri rilevati a istanti precedenti del campione prelevato al tempo t , e per questo motivo vengono denominate carte con memoria. Alcuni lavori sulle carte CUSUM sono stati proposti da Hawkins (1981), Howell (1981) e Chang e Gan (1995). Nei modelli del controllo statistico del processo, si assume che la media stia entro una regione target con un certa variabilità. Il controllo della varianza del processo è molto importante per assicurare la qualità omogenea tra i prodotti. Quando il processo risulta instabile nel tempo, la componente di variabilità è composta essenzialmente da due principali fattori: le cause comuni e le cause speciali. Il metodo classico per le carte di controllo propone di eliminare queste cause speciali di variabilità che si riferiscono principalmente a differenze tra gli operatori, strumenti, materiali, oppure giorni della settimana. Quando esse vengono eliminate, la variabilità residua è dovuta sostanzialmente alle cause comuni, quali, per esempio, cambiamenti ambientali (di solito troppo onerosi e complicati da tenere sotto controllo). Eliminando le cause speciali, il processo torna quindi ad essere in controllo entro la sua variabilità naturale. Sfortunatamente, in alcune situazioni pratiche queste cause non possono essere in principio eliminate o eventualmente ridotte in modo da essere considerate come fattore trascurabile. In presenza di componenti di variabilità tra lotti di prodotti, campioni, oppure simili fonti di varianza, le usuali carte standard \bar{X} e S falliscono, producendo un inaccettabile numero di falsi allarmi. Per ovviare al problema, si è pensato in principio di incrementare i limiti di controllo per una costante arbitraria, ma

così facendo, quando il processo si spostava anche di parecchio dalla sua media, le carte di controllo non erano in grado di rilevare nessun fuori controllo, poichè i limiti erano grandi più del dovuto. Inoltre, quando si voleva scoprire la natura di queste cause, l'uso di un campione statistico che misurasse l'intera variabilità del processo non si rilevava un buon metodo, in quanto un segnale di fuori controllo poteva indicare la presenza di una qualsiasi causa, senza sapere però a cosa potesse riferirsi. In questi casi l'approccio migliore per il controllo della qualità si basa, dopo aver decomposto la varianza globale del processo, sullo studio delle singole componenti di variabilità.

1.2. Esempio reale

Per il controllo sulle componenti di variazione, prendiamo spunto da un caso reale riguardante il processo di realizzazione di circuiti integrati (vedere Chang e Gan, 1994). Nell'assemblaggio delle schede di memoria, una lega deve essere stampata in una basetta (chiamata PCB) prima che i circuiti integrati (IC) vengano montati su essa. Lo spessore della lega è critica nel determinare la compattezza degli IC sulla basetta, quindi è desiderabile che nei PCB ci sia uno spessore omogeneo della lega. La pratica usale prevede di rilevare le misure dello spessore su cinque differenti postazioni dove vengono montati i PCB, e per ciascuna di esse vengono prese due misure dello spessore. Una grande variabilità dentro una basetta PCB può essere dovuta allo stampino non pulito correttamente, mentre una varianza elevata tra le postazioni può essere dovuta a dei blocchi di supporto non allineati. Quindi, è più significativo considerare il controllo della varianza all'interno e tra le postazioni separatamente, piuttosto che considerare un'unica componente di variazione.

Lo spessore della lega, riguardante gli IC, ha media μ e deviazione standard σ_t , la quale, sotto l'assunzione che gli effetti lotto e degli errori interni al lotto siano indipendenti, può essere espressa nel modo seguente:

$$\sigma_t^2 = \sigma_b^2 + \sigma^2,$$

dove σ_b^2 rappresenta la componente di varianza tra i lotti, e σ^2 la varianza delle misure prese all'interno del lotto. I principali motivi che ci spingono a considerare diverse componenti di variazione sono i seguenti:

- 1) Ciascuna componente corrisponde a delle particolari cause assegnabili (cause speciali). Se, per esempio, si presenta un sostanziale incremento della varianza all'interno dei lotti, molto probabilmente essa sarà dovuta a un malfunzionamento degli strumenti, piuttosto che alla diversità di operatori.
- 2) Se la variabilità totale σ_i^2 dipendesse solo da una componente (vale a dire σ_b^2 o σ^2 uguale a 0), sarebbe possibile valutare la qualità del processo su un insieme di valori piuttosto limitati, con riduzioni di costo e fatica.

1.3. Il modello

Di seguito descriviamo il modello usando come esempio il processo riguardante i PCB. Denotando con X lo spessore della lega, il monitoraggio avviene eseguendo delle misurazioni prese su r postazioni dove i PCB vengono montati. Per ciascuna di queste postazioni, vengono prese n misure dello spessore. Le misurazioni seguono le analisi degli effetti casuali del seguente modello di varianza:

$$X_{ij} = \mu + L_i + \varepsilon_{ij}, \quad \text{per } i = 1, 2, \dots, r; \quad j = 1, 2, \dots, n,$$

dove μ rappresenta la media del processo, L_i è l'effetto casuale dell' i -esima postazione, e ε_{ij} l'effetto casuale della j -esima misura prelevata dall' i -esima postazione. Per semplicità, si assume che le componenti di varianza sono indipendenti e normalmente distribuite con media 0 e deviazione standard σ_b e σ , che rispettivamente rappresentano la componente di variabilità tra i lotti e quella delle misurazione presa all' i -esima postazione. In definitiva assumiamo $L_i \sim N(0, \sigma_b^2)$ e $\varepsilon_{ij} \sim N(0, \sigma^2)$.

1.4. Le sequenze di controllo

Avendo definito nel modello i parametri μ , σ_b , σ , le strategie di controllo discusse in questo articolo si basano sullo studio delle tre sequenze $\{\bar{X}_k\}$, $\{\hat{\sigma}_{b(k)}^2\}$, $\{\hat{\sigma}_k^2\}$, dove il pedice indica il k -esimo campione. Ciascun campione è formato da r postazioni (lotti), nelle quali vengono prese in considerazione n misure dello spessore di ossido.

Il controllo del parametro μ viene costruito partendo dalle singole medie calcolate all'interno di un determinato lotto. Avendo a disposizione un numero di lotti pari a r , la stima risulta:

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}, \text{ per } i = 1, \dots, r. \text{ L'indice } i \text{ rappresenta il lotto } i\text{-esimo.}$$

La sequenza di controllo $\{\bar{X}_k\}$ si determina calcolando la media totale dei lotti del k -esimo campione nel modo seguente:

$$\bar{X}_{..} = \frac{1}{rn} \sum_{i=1}^r \sum_{j=1}^n X_{ij}.$$

Quindi ciascuna sequenza $\{\bar{X}_k\}$ rappresenta la media delle medie di ciascun campione. Sotto il modello ipotizzato, abbiamo inoltre che:

$$\bar{X}_i = \mu + Li + \frac{1}{n} \sum_{j=1}^n \varepsilon_{ij},$$

dove le \bar{X}_i rappresentano una sequenza di variabili casuali indipendenti e identicamente distribuite con media pari a μ e varianza definita come $\hat{\sigma}_*^2 = \hat{\sigma}_b^2 + \hat{\sigma}^2/n$, che rappresenta la variabilità tra i lotti all'interno del campione, poichè sotto il modello ipotizzato si ha:

$$E(\bar{X}_i) = \mu + E(Li) + \frac{1}{n} \sum_{j=1}^n E(\varepsilon_{ij}) = \mu,$$

$$\text{Var}(\bar{X}_i) = \text{Var}(Li) + \frac{1}{n^2} n \text{Var}(\varepsilon_{ij}) = \sigma_b^2 + \frac{1}{n} \sigma^2.$$

Il controllo del parametro σ^2 avviene tramite la sequenza $\{\hat{\sigma}_k^2\}$, che rappresenta la stima della dispersione interna ai lotti del k -esimo campione. Il calcolo della sequenza $\{\hat{\sigma}_k^2\}$ diventa:

$$\{\hat{\sigma}_k\} = \hat{\sigma}^2 = \frac{1}{r(n-1)} \sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2,$$

dove le differenze tra le singole osservazioni e le medie delle relative postazioni misurano la dispersione interna a un lotto.

Lo stimatore non distorto della sequenza $\{\sigma_{b(k)}^2\}$, volta a controllare σ_b^2 , è dato invece da:

$$\hat{\sigma}_b^2 = \hat{\sigma}_*^2 - \hat{\sigma}^2 / n,$$

$$\text{con } \hat{\sigma}_*^2 = \frac{1}{r-1} \sum_{i=1}^r (\bar{X}_i - \bar{X}_{..})^2.$$

Quest'ultima componente è espressa come differenza tra la variabilità dei lotti di un campione e la variabilità interna al lotto.

CAPITOLO 2

Disegno delle componenti di varianza

2.1. Carte di controllo Shewhart

Nelle carte di controllo Shewhart, le sequenze $\{\bar{X}_k\}$, $\{\hat{\sigma}_{b(k)}^2\}$, $\{\hat{\sigma}_k^2\}$ discusse al paragrafo 1.4, fanno parte delle statistiche poste sotto osservazione. Queste carte vengono denominate carte senza memoria, in quanto una singola statistica rilevata in un determinato tempo tiene in considerazione solo i dati rilevati nel campione prelevati in quel istante. Sono strettamente legate sulla distribuzione dei dati, in quanto i limiti vengono determinati in base a questa. In generale, costruendo una carta bilaterale, si è soliti fare riferimento ai tre limiti seguenti:

- UCL: costruito partendo dal quantile $1 - \alpha/2$ della distribuzione della statistica sotto osservazione. Il superamento di esso indica presumibilmente che i parametri della distribuzione dei dati sono diversi da quanto ipotizzati.
- CL: linea centrale, indica il quantile 0.5 della distribuzione. Rappresenta la centralità dei dati.
- LCL: costruito partendo dal quantile $\alpha/2$ della distribuzione della statistica.

La carta Shewhart è uno strumento diffusamente utilizzato perchè è di facile interpretazione. Quando si decide di sorvegliare la varianza tramite questa carta, se un punto della statistica sarà superiore al limite UCL, questo indicherà sostanzialmente un deterioramento della qualità del processo, al contrario se il punto cadrà al di sotto del limite LCL, questo può significare un miglioramento della qualità del processo poiché la variabilità dal valore medio della statistica si è ridotta. Quando si decide di sorvegliare la media, anche un punto che cadrà al di sotto del limite inferiore, darà l'indicazione di uno spostamento del valore medio nella regione di rifiuto.

2.2. Disegno della carta Shewhart per il controllo di $\hat{\sigma}^2$

Per il monitoraggio della componente σ^2 , i limiti vengono calcolati sfruttando la distribuzione del rapporto $\hat{\sigma}^2 / \sigma^2$. La variabile casuale $r(n-1)\hat{\sigma}^2 / \sigma^2 \sim \chi_{r(n-1)}^2$, dove χ_ν^2 rappresenta la distribuzione chi-quadro con ν gradi di libertà. La media e la varianza di $\hat{\sigma}^2$ sono rispettivamente:

$$E(\hat{\sigma}^2) = \sigma^2,$$

$$Var(\hat{\sigma}^2) = 2\sigma^4 / r(n-1).$$

Così, i limiti di controllo per $\hat{\sigma}^2$ possono essere determinati facilmente per un dato tasso di falso allarme α .

Dato il tasso di falso allarme α , i limiti di controllo per la carta Shewhart si ottengono nel modo seguente:

$$UCL = \frac{\sigma^2 \chi_{1-\alpha/2}^2(r(n-1))}{r(n-1)}, \quad CL = \frac{\sigma^2 \chi_{0.5}^2(r(n-1))}{r(n-1)}, \quad LCL = \frac{\sigma^2 \chi_{\alpha/2}^2(r(n-1))}{r(n-1)},$$

dove σ^2 può essere nota o stimata usando dati da un processo in controllo, α è il livello di significatività osservato, e $\chi_\gamma^2(\nu)$ indica il quantile γ della distribuzione chi-quadro con ν gradi di libertà.

2.3. Disegno della carta Shewhart per il controllo di $\hat{\sigma}_b^2$

Il disegno della carta di controllo per σ_b^2 risulta più laborioso, in quanto essa è composta da due componenti al suo interno: quella erratica che fa riferimento alla varianza entro le postazioni, e la variabilità tra le medie delle postazioni. Per l'implementazione del disegno, occorre partire dal calcolo della distribuzione di $\hat{\sigma}_b^2$, la quale viene espressa come differenza di due variabili casuali ciascuna proporzionale a una variabile chi quadro:

$$\hat{\sigma}_b^2 = S - T / n, \text{ dove}$$

$$S = \hat{\sigma}_*^2 \text{ con } (r-1)S / \sigma_*^2 \sim \chi_{r-1}^2, \text{ e}$$

$$T = \hat{\sigma}^2 \text{ con } r(n-1)T / \sigma^2 \sim \chi_{r(n-1)}^2.$$

Le funzioni di densità di probabilità per S e T saranno rappresentate rispettivamente da $f_{\sigma_*,u}(s)$ e $f_{\sigma,v}(t)$, con $u = (r-1)/2$ e $v = r(n-1)/2$, dove i deponenti rappresentano la radice quadrata della vera varianza e 0.5 volte i gradi di libertà. Dato che S e T sono indipendenti e hanno funzione di densità continua, la loro funzione di densità congiunta è continua ed è data dal prodotto delle loro funzioni di densità. Mood, Graybill, e Boes (1974, p.185) danno la funzione di densità della differenza tra 2 variabili casuali, e sostituendo nelle loro espressioni le funzioni $f_{\sigma_*,u}(s)$ e $f_{\sigma,v}(t)$ si ottiene la funzione di densità per $Y = \widehat{\sigma}_b^2$ data da:

$$\frac{d}{dy} P(Y \leq y) = \int_{x=-\infty}^{\infty} f_{\sigma_*,u}(x+y) n f_{\sigma,v}(nx) dx.$$

Integrando ulteriormente, si ottiene la funzione di ripartizione di Y , denominata CDF, data da:

$$P(Y \leq y) = \int_{w=-\infty}^y \int_{x=-\infty}^{\infty} f_{\sigma_*,u}(x+w) n f_{\sigma,v}(nx) dx dw.$$

La media e la varianza di $\widehat{\sigma}_b^2$ sono rispettivamente:

$$E(\widehat{\sigma}_b^2) = \sigma_b^2,$$

$$Var(\widehat{\sigma}_b^2) = 2\sigma_*^4 / (r-1) + 2\sigma^4 / (n^2 r(n-1)).$$

Il calcolo della CDF permette di ricavare i vari quantili per $\widehat{\sigma}_b^2$. Per non dover necessariamente ricorrere al calcolo di quest'ultima, e per permetterne un uso scorrevole, i quantili sono stati ottenuti con una stima molto precisa mediante modelli di regressione. Quindi per valori di α 0.001, 0.002, 0.005, è possibile ottenere i coefficienti del modello di regressione per determinare i limiti di controllo (vedere Chang e Gan, 1994) per diversi valori di r ed n e del rapporto σ_b / σ compresi tra [0.05, 3]. La linea centrale è calcolata prendendo il 0.5 quantile di $\widehat{\sigma}_b^2$.

Per determinare i limiti di controllo della carta Shewhart occorre effettuare i seguenti passi:

- 1) Dopo aver fissato r ed n , bisogna calcolare il valore $z = \sigma_b / \sigma$ quando σ_b^2 e σ^2 sono in controllo.
- 2) Dopo aver determinato il valore z , il limite UCL, corrispondente al quantile $(1-\alpha/2)$, si ottiene dalla seguente equazione:

$$UCL = \sigma^2 (\beta_0 + \beta_1 z + \beta_2 z^2),$$

i cui coefficienti sono riportati nelle tabelle nell'articolo di Chang e Gan (1994, da pagina 296). Quando r è un numero pari, i limiti di controllo esatti non possono essere ottenuti perchè la formula esatta della CDF per $\hat{\sigma}_b^2$ non può essere calcolata. Perciò il limite viene approssimato usando una media degli UCL calcolati per $(r-1)$ e $(r+1)$.

Analogamente il limite LCL, corrispondente al quantile $\alpha/2$, si ottiene dalla seguente equazione:

$$LCL = \sigma^2(\beta_0 + \beta_1 z + \beta_2 z^2),$$

e valgono le stesse considerazioni fatte per il limite UCL. I coefficienti sono riportati nell'articolo di Chang e Gan (1994, da pagina 299).

- 3) Si determina la linea centrale utilizzando i coefficienti appropriati, secondo la formula seguente:

$$CL = \sigma^2(\beta_0 + \beta_1 z + \beta_2 z^2).$$

Anche in questo caso se r è un numero pari, il limite centrale viene approssimato usando la media tra le LC calcolate per $(r-1)$ e $(r+1)$. I coefficienti sono riportati nell'articolo di Chang e Gan (1994, pagina 303).

2.4. Disegno della carta Shewhart per il controllo di μ

Il controllo della media del processo si basa sulla sequenza $\{\bar{X}_k\}$, che rappresenta per il k -esimo campione la media delle medie delle postazioni. La sequenza di controllo $\{\bar{X}_k\}$ esprimibile come:

$$\{\bar{X}_k\} = \bar{X}_{..} = \mu + \frac{1}{r} \sum_{l=1}^r Li + \frac{1}{rn} \sum_{i=1}^r \sum_{j=1}^n \varepsilon_{ij},$$

ha valore medio pari a μ , perchè

$$E\{\bar{X}_k\} = E(\bar{X}_{..}) = \mu,$$

e varianza pari a

$$Var\{\bar{X}_k\} = \frac{1}{r^2} r \sigma_b^2 + \frac{1}{r^2 n^2} r n \sigma^2 = \frac{1}{r} \sigma_b^2 + \frac{1}{rn} \sigma^2.$$

Poiché assumiamo la normalità delle osservazioni, si ottengono i seguenti limiti di controllo:

$$UCL = \mu + Z_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_b^2}{r} + \frac{\bar{\sigma}^2}{rn}}, CL = \hat{\mu} = \mu, LCL = \mu - Z_{1-\alpha/2} \sqrt{\frac{\hat{\sigma}_b^2}{r} + \frac{\bar{\sigma}^2}{rn}},$$

dove Z_q denota il quantile q -esimo della distribuzione normale standard.

2.5. Carte di controllo CUSUM

Le carte CUSUM vengono dette carte senza memoria in quanto tengono conto del peso delle statistiche rilevate a istanti precedenti dal tempo t . Queste carte sono state inizialmente proposte da Page (1954), e successivamente da diversi autori: Ewan (1963), Page(1961), Gan(1991), Lucas (1976), Hawkins (1981,1993a), Yaschin (1985), Woodall e Adams(1993).Nello studio delle componenti di varianza, la carta CUSUM per la variabilità viene definita come un unico schema che raccoglie le deviazioni dalla costante k , la quale dipende strettamente dallo scostamento che si vuole individuare. Prima vengono definite le regioni di accettabilità e di rifiuto determinando così il valore k (si veda la figura 2.1), poi si procede col costruire la statistica vera e propria. Se la variabilità è maggiore della costante k , la statistica accumula lo scarto tra σ e k mantenendo questa informazione anche per gli istanti successivi (carte con memoria). Dopo aver determinato il valore k , rimane da scegliere il limite di controllo h (denominata soglia). Di solito la scelta di questa viene suggerita in modo da ottenere buone prestazioni in termini di ARL per quanto riguarda lo schema (lo studio di questa verrà affrontato successivamente). Va però detto che la scelta di h dipende dalla scelta di k , in quanto a parità di ARL, a un valore di k più basso, quindi a una regione di rifiuto molto vicina a quella di accettabilità, corrisponde un valore di soglia più alto. Perciò la scelta di k e h è cruciale per determinare la sensibilità dello schema. Negli esempi apportati, il valore h viene ricavato tramite simulazione.

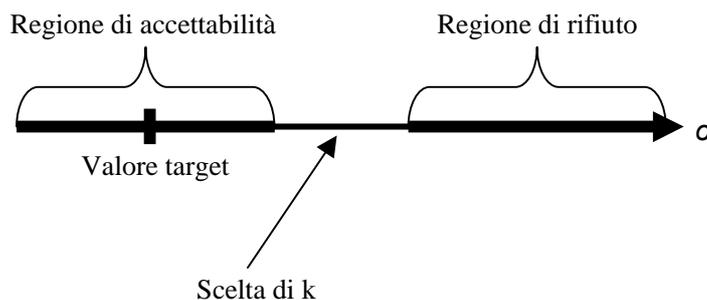


Figura 2.1: La scelta del valore k è posta tra la regione di accettabilità e quella di rifiuto.

2.6. Disegno della carta CUSUM per il controllo di $\hat{\sigma}^2$

Il monitoraggio di σ si basa sulle sequenze di controllo $\{\sigma_k^2\}$ le quali, per il campione k -esimo, tengono conto della dispersione totale entro i lotti. Dopo aver definito le regioni di accettabilità e di rifiuto per σ date rispettivamente da $\sigma \leq \sigma_0$ e $\sigma \geq \sigma_1$, con $(\sigma_0 \leq \sigma_1)$, lo schema della statistica CUSUM è definito come segue:

$$S_0 = 0, \quad S_k = \max[S_{k-1} + (\hat{\sigma}_k^2 - k), 0],$$

dove il pedice k indica la statistica del k -esimo campione, per $k=1,2,\dots$

La statistica S_k sarà maggiore di 0 quando le deviazioni $(\hat{\sigma}_k^2 - k) > 0$, indicazione che la variabilità del processo si è spostata dentro un regione inaccettabile. Il valore di riferimento k si calcola nel modo seguente:

$$k = k_\sigma = \frac{2 \ln(\sigma_1 / \sigma_0)}{\sigma_0^{-2} - \sigma_1^{-2}}.$$

2.7. Disegno della carta CUSUM per il controllo di $\hat{\sigma}_b^2$

Per questa componente di varianza, il disegno dello schema CUSUM è piuttosto complesso, principalmente per la presenza di un parametro di disturbo. Nel problema di controllo di σ_b , il parametro di disturbo è σ . Questo parametro rappresenta la parte di σ_*^2 attribuibile alla dispersione all'interno di un lotto. Denotiamo con $\eta = \sigma^2 / n$ e con σ_{b0} e σ_{b1} rispettivamente il limite superiore della regione di accettabilità e il limite inferiore della regione di rifiuto, con $\sigma_{b0} < \sigma_{b1}$. Trattiamo in primo luogo il caso in cui η sia noto.

1) *Caso η noto*: in questo caso la sequenza di controllo per il monitoraggio di σ_b^2 si basa sulla sequenza degli stimatori non distorti $\{\sigma_{*(k)}^2 - \eta\}$. Lo schema CUSUM diventa:

$$S_0 = 0 \quad S_k = \max[S_{k-1} + \hat{\sigma}_{*(k)}^2 - \eta - k_b(\eta)],$$

$$\text{con } k_b(\eta) = -\eta + \frac{\ln[(\sigma_{b1}^2 + \eta)/(\sigma_{b0}^2 + \eta)]}{(\sigma_{b0}^2 + \eta)^{-1} - (\sigma_{b1}^2 + \eta)^{-1}}.$$

Un deterioramento della qualità si avrà quando la statistica supererà il valore di soglia h .

2) *Caso η non noto*: nella maggior parte delle situazioni pratiche, il parametro η non è noto, anche se un'analisi dei dati in nostro possesso potrebbe specificare un possibile campo di variazione dei suoi possibili valori. In questo caso il controllo della componente σ_b si basa sulla sequenza $\{\hat{\sigma}_{*(k)}^2 - \hat{\sigma}_k^2\}$, ovvero $\{\hat{\sigma}_{b(k)}^2\}$. Lo schema CUSUM diventa:

$$S_0 = 0 \quad S_k = \max[S_{k-1} + \hat{\sigma}_{b(k)}^2 - k_b(\tilde{\eta})].$$

Il valore $\tilde{\eta}$ è un valore ragionevole per η posto pari a $\hat{\sigma}^2 / n$ e il valore $k_b(\tilde{\eta})$ viene calcolato usando la formula precedente con η noto. Anche qui un deterioramento della qualità si avrà quando la statistica S_k supererà il valore di soglia h .

2.8. Disegno della carta CUSUM per il controllo di μ

L'implementazione dello schema di controllo per μ si basa sul definire la coppia di schemi denotati con $\{S_k^+\}$ e $\{S_k^-\}$, le statistiche che accumulano deviazioni dal valore target μ_0 superiori e inferiori al valore k , rispettivamente. Questi vengono denominati in genere CUSUM unilaterale superiore, e CUSUM unilaterale inferiore. Lo schema può essere definito nella seguente maniera:

$$S_0^+ = 0, S_k^+ = \max[S_{k-1}^+ + (\bar{X}_k - k^+), 0],$$

$$S_0^- = 0, S_k^- = \max[S_{k-1}^- + (k^- - \bar{X}_k), 0].$$

Il valore di riferimento superiore k^+ viene scelto tipicamente tra il limite superiore della regione di accettabilità, e il limite inferiore della regione di rifiuto, come accadeva per le altre componenti, e può essere espresso come:

$$k^+ = \mu_0 + \left| \frac{\mu_1 - \mu_0}{2} \right|,$$

dove μ_0 è il valore target della media, e μ_1 rappresenta il livello medio del processo instabile.

Il valore di riferimento inferiore k^- viene invece scelto tra il limite inferiore della regione di accettabilità, e il limite superiore della regione di rifiuto, e può essere espresso come:

$$k^- = \mu_0 - \left| \frac{\mu_1 - \mu_0}{2} \right|.$$

In generale il valore $\left| \frac{\mu_1 - \mu_0}{2} \right|$ viene denotato con K , il quale è definito come :

$$K = \frac{1}{\sigma} \left| \frac{\mu_1 - \mu_0}{2} \right| \sigma = \frac{\delta}{2} \sigma = k \sigma.$$

La scelta della soglia per la media viene generalmente definita come:

$$H = h \sigma,$$

dove h è un parametro strettamente dipendente dalla scelta del valore k , semiampiezza del salto standardizzato:

$$k = \frac{\delta}{2} = \left| \frac{\mu_1 - \mu_0}{2} \right| \frac{1}{\sigma},$$

dove σ rappresenta la stima della deviazione standard della media. Sono stati fatti molti studi riguardo la scelta del parametro h , e mediante apposite tabelle (Hawkins, 1992,1993a), per determinati valori k , è possibile ricavarne il valore.

CAPITOLO 3

Costruzione delle carte di controllo

Questo capitolo illustrerà come applicare i procedimenti affrontati nei paragrafi precedenti nel disegno delle carte di controllo, e servirà per fare un confronto tra le performance delle due carte, Shewhart e CUSUM, rispetto a cambiamenti nelle singole componenti di variazione e nella media. Per l'applicazione consideriamo l'esempio fatto nei primi paragrafi relativo alla costruzione dei circuiti IC, dove viene ispezionato lo strato di spessore della lega prendendo $r = 5$ locazioni dove vengono montati i PCB, con $n = 2$ misure per ciascuna postazione. Da un'analisi condotta su 500 campioni di dati ritenuti in controllo, le stime per σ e σ_b sono rispettivamente 7.135 e 7.014, mentre il valore medio μ è pari a 35.

3.1. Costruzione per le carte Shewhart

3.1.1. Componente $\hat{\sigma}^2$

Considerando una carta Shewhart bilaterale con un valore $\alpha = 0.005$, i limiti di controllo per la componente σ risultano essere $UCL = 187.2$, $CL = 44.3$, ed $LCL = 3.13$. L'ARL corrispondente calcolata come $1/\alpha$, è pari a 200. Quindi se il processo è in controllo ci aspettiamo di rilevare un falso allarme mediamente una volta ogni duecento campioni.

Con le formule date precedentemente al punto 2.1, i limiti vengono calcolati nel seguente modo:

$$UCL = \frac{7.135^2(18.90738)}{5(2-1)} = 187.2, \quad \text{con} \quad \chi_{1-0.005/2}^2(5(2-1)) = 18.38561$$

$$CL = \frac{7.135^2(4.35146)}{5(2-1)} = 44.3, \quad \text{con} \quad \chi_{0.5}^2(5(2-1)) = 4.35146.$$

$$LCL = \frac{7.135^2(0.3074818)}{5(2-1)} = 3.13, \quad \text{con} \quad \chi_{0.005}^2(5(2-1)) = 0.3074818$$

3.1.2. Componente $\hat{\sigma}_b^2$

Per il controllo della componente σ_b^2 , consideriamo uno schema unilaterale con $\alpha = 0.002$. Quindi l'ARL in controllo pari a $1/\alpha$ è uguale a 500. Usando i tre passi per la costruzione dei limiti di controllo descritti al paragrafo 2.3, otteniamo i seguenti risultati:

Passo 1: dati i valori $r = 5$ ed $n = 2$, si trova $z = 7.014/7.135 = 0.983$.

Passo 2: dopo aver fissato i valori di r ed n , con $\alpha = 0.002$, i coefficienti del modello di regressione per determinare UCL sono $\beta_0 = 1.7461$, $\beta_1 = -0.094615$, $\beta_2 = 4.2496$ (Chang e Gan, 1994, pagina 297). Con z pari a 0.983, e $\sigma^2 = 50.908$, il valore UCL trovato risulta pari a 293.2, poichè:

$$UCL = 50.908(1.7461 - (0.094615)z + (4.2496)z^2) = 293.2.$$

Passo 3: per la determinazione della linea centrale, i coefficienti del modello di regressione risultano $\beta_0 = -0.0072305$, $\beta_1 = -0.05503$, $\beta_2 = 0.85025$ (Chang e Gan, 1994, pagina 303). Con z pari a 0.983, e $\sigma^2 = 50.908$, il valore CL trovato risulta pari a 38.7, poichè:

$$CL = 50.908(-0.0072305 - (0.005503)z + (0.85025)z^2) = 38.7$$

3.1.3. Componente μ

Infine, per il controllo della componente μ consideriamo lo schema bilaterale con $\alpha = 0.005$. Lo schema è costruito per ottenere un ARL pari a 200. Il limite centrale per la carta di controllo è pari a 35 (livello del processo noto a priori). La stima della varianza per \bar{X}_r risulta 14.93006, poichè :

$$Var\{\bar{X}_r\} = \frac{1}{5}7.014^2 + \frac{1}{10}7.135^2 = 14.93006.$$

Con $\alpha = 0.005$, e $Z_{1-0.005/2} = 2.807034$, i limiti ottenuti superiore e inferiore sono:

$$UCL = 35 + 2.8\sqrt{14.93006} = 45.84622,$$

$$LCL = 35 - 2.8\sqrt{14.93006} = 24.15378.$$

3.2. Costruzione per le carte CUSUM

Per la costruzione delle carte CUSUM siamo interessati, sia per media che per la variabilità, al disegno di schemi che siano sensibili a un cambiamento di ampiezza pari a 1.5, 1, e 0.5 volte lo standard error della componente analizzata.

3.2.1. Componente $\hat{\sigma}^2$

Per quanto riguarda il controllo della componente σ , definiamo per prima cosa le regioni di accettabilità ed i rifiuto. Le regioni di rifiuto sono, per i salti pari a 1.5, 1, 0.5 volte lo standard error di σ^2 , $\sigma > \sigma_1 = 9.96012$, $\sigma > \sigma_1 = 9.1162$, $\sigma > \sigma_1 = 8.18577$..

Poichè lo standard error di $\hat{\sigma}^2$ risulta

$$s.e(\hat{\sigma}^2) = \sqrt{\frac{2(7.135)^4}{5(2-1)}} = 32.19719,$$

i salti proporzionali rispetto a σ si calcolano nella seguente maniera:

- salto di 1.5 $s.e(\hat{\sigma}^2)$: $\sqrt{(7.135^2 + 1.5s.e(\hat{\sigma}^2))} = \sqrt{(7.135^2 + 48.29578)} = 9.96012$.
- salto di 1 $s.e(\hat{\sigma}^2)$: $\sqrt{(7.135^2 + 1s.e(\hat{\sigma}^2))} = \sqrt{(7.135^2 + 32.19719)} = 9.1162$.
- salto di 0.5 $s.e(\hat{\sigma}^2)$: $\sqrt{(7.135^2 + 0.5s.e(\hat{\sigma}^2))} = \sqrt{(7.135^2 + 16.09859)} = 8.18577$.

I valori di riferimento k vengono calcolati nella maniera seguente:

- per il salto di 1.5 $s.e(\hat{\sigma}^2)$: $k = k_\sigma = \frac{2 \ln(9.96012 / 7.135)}{7.135^{-2} - 9.96012^{-2}} = 69.76442$,
- per il salto di 1 $s.e(\hat{\sigma}^2)$: $k = k_\sigma = \frac{2 \ln(9.1162 / 7.135)}{7.135^{-2} - 9.1162^{-2}} = 64.3978$,
- per il salto di 0.5 $s.e(\hat{\sigma}^2)$: $k = k_\sigma = \frac{2 \ln(8.18577 / 7.135)}{7.135^{-2} - 8.18577^{-2}} = 58.22214$.

Dovendo fare un confronto con la rispettiva carta Shewhart per questa componente, la soglie scelte a garantire un ARL in controllo pari a 200 per i tre salti risultano approssimativamente:

- per il salto di 1.5 $s.e(\hat{\sigma}^2)$ pari a 136.4,
- per il salto di 1 $s.e(\hat{\sigma}^2)$ pari a 162,
- per il salto di 0.5 $s.e(\hat{\sigma}^2)$ pari a 213.3.

Si noti che quando uno schema è volto a individuare scostamenti piccoli per una generica componente, esso avrà un valore k più basso rispetto a schemi sensibili per scostamenti di dimensioni maggiori. Di conseguenza, a parità di ARL in controllo fissata, la soglia risultante sarà più alta rispetto agli altri schemi .

3.2.2. Componente $\hat{\sigma}_b^2$

Per il controllo di σ_b , procedendo analogamente come per l'altra componente, le regioni di rifiuto diventano $\sigma_b > \sigma_{b1} = 11.48805$, $\sigma_b > \sigma_{b1} = 10,217$, $\sigma_b > \sigma_{b1} = 10,217$, per i salti rispettivi di 1.5, 1 e 0.5 volte lo standard error di $\hat{\sigma}_b^2$. Con il valore di $\hat{\sigma}_*^2 = 7.014^2 + 7.135^2 / 2 = 74.65031$, la deviazione standard risultante per $\hat{\sigma}_b^2$ risulta

$$s.e(\hat{\sigma}_b^2) = \sqrt{\frac{2(74.65031)^2}{(5-1)} + \frac{2(7.135^4)}{2^2(5(2-1))}} = 55.18604,$$

e i salti proporzionali rispetto a σ_b risultano :

- salto di 1.5 $s.e(\hat{\sigma}_b^2)$: $\sqrt{(7.014^2 + 1.5s.e(\hat{\sigma}_b^2))} = \sqrt{(7.014^2 + 82.77906)} = 11.48805$,
- salto di 1 $s.e(\hat{\sigma}_b^2)$: $\sqrt{(7.014^2 + 1s.e(\hat{\sigma}_b^2))} = \sqrt{(7.014^2 + 55.18604)} = 10.21676$,
- salto di 0.5 $s.e(\hat{\sigma}_b^2)$: $\sqrt{(7.014^2 + 0.5s.e(\hat{\sigma}_b^2))} = \sqrt{(7.014^2 + 27.59302)} = 8.762946$.

I valori di riferimento k vengono calcolati nella seguente maniera:

- per il salto di 1.5 $s.e(\hat{\sigma}_b^2)$:

$$k_w(\tilde{\eta}) = -74.6503 + \frac{\ln[(11.48805^2 + 74.6503)/(7.014^2 + 74.6503)]}{(7.014^2 + 74.6503)^{-1} - (11.48805^2 + 74.6503)^{-1}} = 80.47862,$$

- per il salto di 1 $s.e(\hat{\sigma}_b^2)$:

$$k_w(\tilde{\eta}) = -74.6503 + \frac{\ln[(10.21676^2 + 74.6503)/(7.014^2 + 74.6503)]}{(7.014^2 + 74.6503)^{-1} - (10.21676^2 + 74.6503)^{-1}} = 71.75,$$

- per il salto di 0.5

$s.e(\hat{\sigma}_b^2)$:

$$k_w(\tilde{\eta}) = -74.6503 + \frac{\ln[(8.762946^2 + 74.6503)/(7.014^2 + 74.6503)]}{(7.014^2 + 74.6503)^{-1} - (8.762946^2 + 74.6503)^{-1}} = 61.55094.$$

Le soglie scelte in modo tale da garantire un ARL in controllo approssimativamente pari a 500 risultano:

- per il salto di 1.5 $s.e(\hat{\sigma}_b^2)$ h pari a 301.55,
- per il salto di 1 $s.e(\hat{\sigma}_b^2)$ h pari a 359,
- per il salto di 0.5 $s.e(\hat{\sigma}_b^2)$ h pari a 485.

3.2.3. Componente μ

Per il controllo sulla media, definiamo la seguenti aree di rifiuto come inaccettabili:

- per il salto di 1.5 $s.e(\mu)$: $|\mu - 35| > 5.795916$,
- per il salto di 1 $s.e(\mu)$: $|\mu - 35| > 3.863944$,
- per il salto di 0.5 $s.e(\mu)$: $|\mu - 35| > 1.931972$.

Poiché la deviazione standard di μ risulta:

$$s.e(\mu) = \sqrt{\frac{7.014^2}{5} + \frac{7.135^2}{2}} = 3.863944,$$

otteniamo i seguenti valori intermedi come limiti di accettabilità:

- per il salto di 1.5 $s.e(\mu)$:

$$k^+ = 35 + \left| \frac{40.79592 - 35}{2} \right| = 37.89796,$$

$$k^- = 35 - \left| \frac{40.79592 - 35}{2} \right| = 32.10204.$$

- per il salto di $1 \text{ s.e}(\mu)$:

$$k^+ = 35 + \left| \frac{38.86394 - 35}{2} \right| = 36.93197,$$

$$k^- = 35 - \left| \frac{38.86394 - 35}{2} \right| = 33.06803,$$

- per il salto di $0.5 \text{ s.e}(\mu)$:

$$k^+ = 35 + \left| \frac{36.93197 - 35}{2} \right| = 35.96599,$$

$$k^- = 35 - \left| \frac{36.93197 - 35}{2} \right| = 34.03401.$$

Tramite le tabelle proposte da Hawkins (1992), che forniscono una tavola di valori k corrispondenti a un valore h per un dato valore di ARL, è possibile ricavare le seguenti soglie:

- per il salto di $1.5 \text{ s.e}(\mu)$, con k pari a 0.75, si trova h pari a 2.9, e la soglia corrispondente 11.28272.
- per il salto di $1 \text{ s.e}(\mu)$, con k pari a 0.5, si trova h pari a 4.2, e la soglia corrispondente 16.22856.
- per il salto di $0.5 \text{ s.e}(\mu)$, con k pari a 0.25, si trova h pari a 6.8, e la soglia corrispondente 26.27482.

Le soglie sono calcolate in modo tale da mantenere un ARL in controllo approssimativamente pari a 200, per uno schema bilaterale.

CAPITOLO 4

Confronti tra le carte di controllo

4.1. Confronto per la componente σ

Per i raffronti tra le due carte, si è scelto di simulare 30 campioni da una distribuzione con parametri sotto controllo posti pari $\sigma = 7.135$, $\sigma_b = 7.014$, $\mu = 35$, e per ogni esempio di aggiungere 50 campioni da una distribuzione la cui variabilità σ viene aumentata di un valore proporzionale alla sua deviazione standard. Per la carta di controllo Shewhart i limiti calcolati al paragrafo 3.1.1. valgono per ogni esempio riportato, mentre per la carta CUSUM, la soglia h sarà cambiata in modo da garantire uno schema sensibile rispetto al rispettivo cambiamento del parametro σ .

Il primo esempio metterà a confronto il comportamento delle due carte con un salto di σ pari a 1.5 volte il suo standard error, vale a dire $\sigma = 9.96012$.

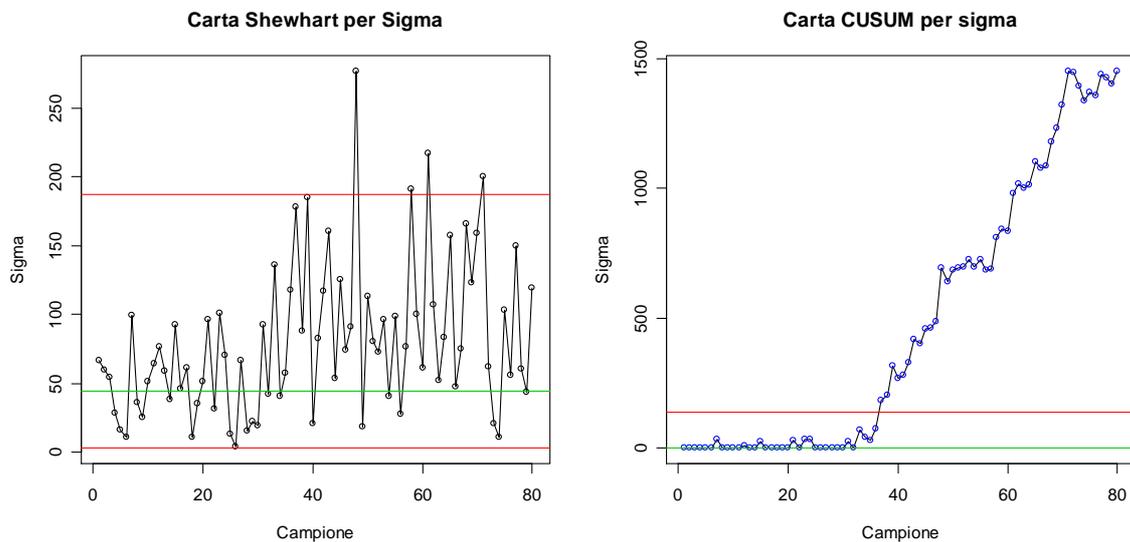


Figura 4.1: Confronto tra la carta Shewhart e CUSUM per σ pari a 9.96. La carta Shewhart è costruita con i limiti $UCL = 187.2$, $CL = 44.3$, $LCL = 3.13$, la carta CUSUM costruita con una soglia pari a 136.4.

Guardando la figura 4.1, si nota che in presenza di salti piuttosto alti della variabilità, come spesso accade, la carta Shewhart si dimostra efficiente. In questo esempio abbiamo che prima del 40° campione una statistica sta a ridosso del limite superiore, e si può notare un salto della centralità dei dati dopo il 30° campione, segnalazione che il livello medio di variabilità è aumentato. La carta CUSUM si comporta bene, segnalando una deriva crescente dopo il 30° campione.

Valutiamo il comportamento delle due carte aggiungendo ai 30 campioni originali, 50 campioni con uno shift in varianza pari a una volta lo standard error di $\hat{\sigma}^2$. In questo caso σ è pari a 9.116217.

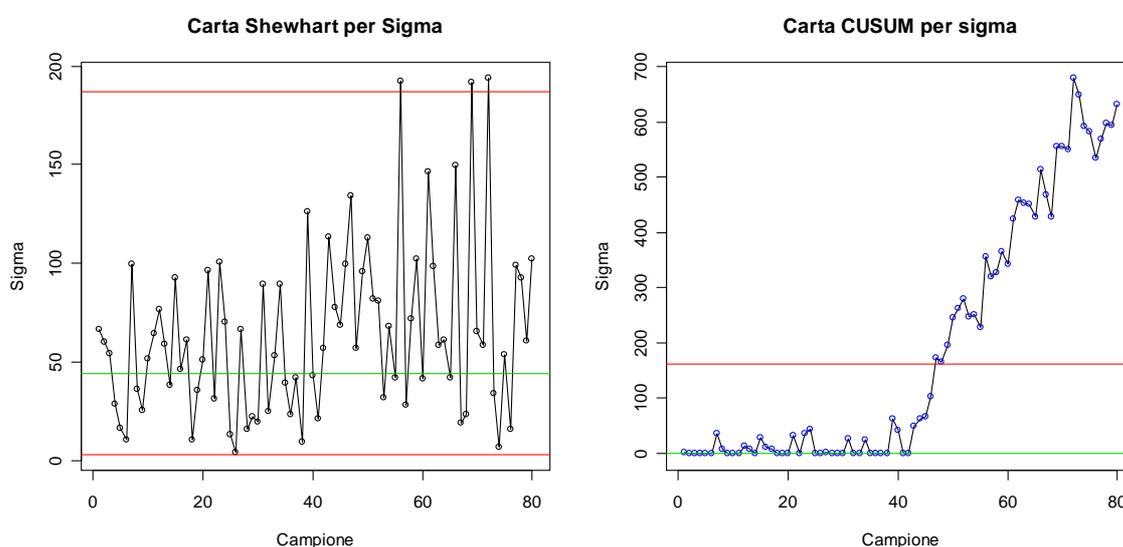


Figura 4.2: Confronto tra la carta Shewhart e CUSUM per σ pari a 9.116. La carta Shewhart è costruita con i limiti $UCL = 187.2$, $CL = 44.3$, $LCL = 3.13$, la carta CUSUM è costruita con una soglia pari a 162.

Entrambe le carte mostrano dei fuori controllo, segnale che il processo non sta lavorando con parametri ritenuti in controllo per la produzione. La carta CUSUM presenta una deriva crescente che si manifesta poco dopo il 40 campione, mentre per la carta Shewhart la prima statistica di fuori controllo avviene all'altezza del 56 campione, più tardi quindi rispetto alla CUSUM. La deriva positiva si presenta poichè dopo che il processo non sta più lavorando con la variabilità σ^2 in controllo, la statistica accumula man mano valori positivi dovuti al cambiamento di questa componente. In questo caso il comportamento delle due carte risulta soddisfacente, anche se la carta con memoria segnala prima il fuori controllo, quindi si comporta meglio.

Come prossimo esempio aggiungiamo ai 30 campioni originali sempre 50 campioni questa volta con uno shift in varianza a 0.5 lo standard error di $\hat{\sigma}^2$. In questo caso σ è pari a 8.18577.

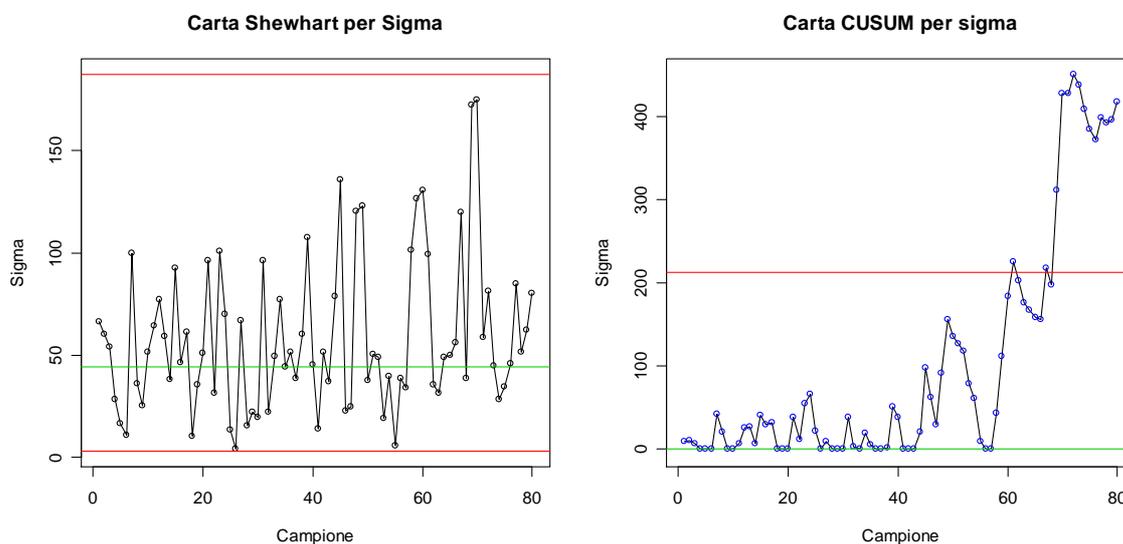


Figura 4.3: Confronto tra la carta Shewhart e CUSUM per σ pari a 8.18. La carta Shewhart è costruita con i limiti $UCL = 187.2$, $CL = 44.3$, $LCL = 3.13$, la carta CUSUM costruita con una soglia pari a 213.3.

Come si può notare dalla figura 4.3, la flessibilità della carta CUSUM permette di rilevare anche salti piccoli in varianza, segnalando dei fuori controllo dopo il 60° campione, anche se in maniera tardiva rispetto all'esempio riportato precedentemente. Per questo piccolo cambiamento la carta Shewhart non segnala nulla di particolare, quindi resta quantomeno difficile valutare se è avvenuto un cambiamento della variabilità (si noti comunque che la statistica raggiunge dei picchi più alti rispetto ai primi 30 campioni).

4.2. Confronto per la componente σ_b

Per lo studio della componente più rilevante, gli esempi seguiranno la falsa riga di quelli precedenti, considerando 30 campioni con i parametri considerati sotto controllo, vale a dire $\sigma = 7.135$, $\sigma_b = 7.014$, $\mu = 35$, e sempre per ognuno 50 campioni con un salto della variabilità σ_b^2 proporzionale al suo standard error. Per la carta di controllo Shewhart i limiti calcolati al paragrafo 3.1.2. valgono per ogni esempio riportato, mentre per la carta CUSUM, la soglia h sarà cambiata in modo da garantire uno schema sensibile rispetto al

rispettivo cambiamento del parametro σ_b . I valori di tali soglie sono recuperabili al paragrafo 3.2.2.

Come primo esempio riportiamo 30 campioni in controllo, con 50 che presentano uno shift in varianza pari a 1.5 volte lo standard error di $\hat{\sigma}_b^2$. Il valore σ_b risulta pari a 11.48805.

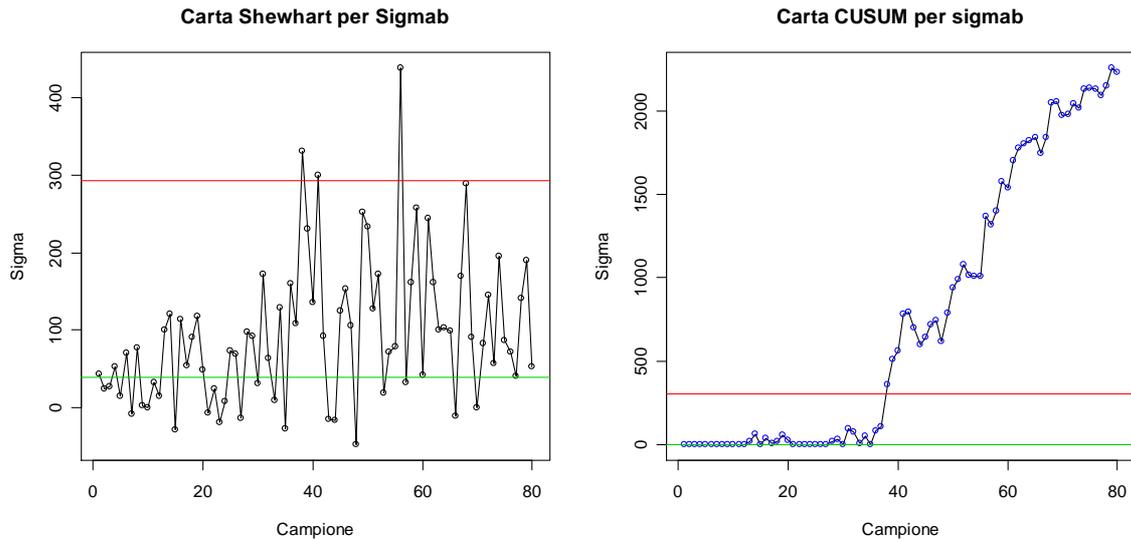


Figura 4.4: Confronto tra la carta Shewhart e CUSUM per σ_b pari a 11.48. La carta Shewhart è costruita con i limiti $UCL = 293.2$, $CL = 38.7$, la carta CUSUM è costruita con una soglia pari a 301.55.

Le due carte si comportano bene rispetto a un cambiamento cospicuo della variabilità σ_b . La deriva crescente della carta CUSUM inizia poco dopo il 30° campione, mentre la carta Shewhart segnala dei fuori controllo subito dopo il cambio in varianza.

Valutiamo il comportamento delle due carte, questa volta prendendo in considerazione uno shift in varianza pari a una volta la deviazione standard di $\hat{\sigma}_b^2$. Abbiamo un valore $\sigma_b = 10.21676$.

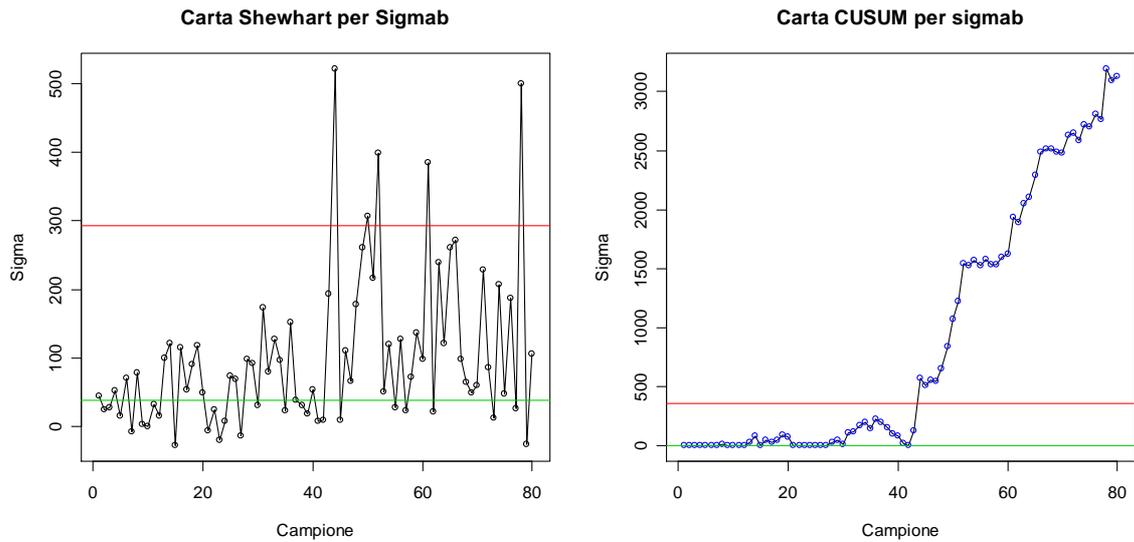


Figura 4.5: Confronto tra la carta Shewhart e CUSUM per σ_b pari a 10.21. La carta Shewhart è costruita con i limiti $UCL = 293.2$, $CL = 38.7$, la carta CUSUM è costruita con una soglia pari a 359.

Guardando la figura 4.5, dopo il 30° campione la carta di CUSUM segnala una deriva positiva crescente, e la carta Shewhart presenta parecchi fuori controllo dopo il 40° campione, risultando così ugualmente efficienti. Anche qui va notato l'aumento della centralità della statistica Shewhart dopo il cambiamento avvenuto in varianza.

Concludiamo il raffronto prendendo in considerazione uno shift in varianza pari a 0.5 volte la deviazione standard di $\hat{\sigma}_b^2$. Abbiamo un valore $\sigma_b = 8.762946$.

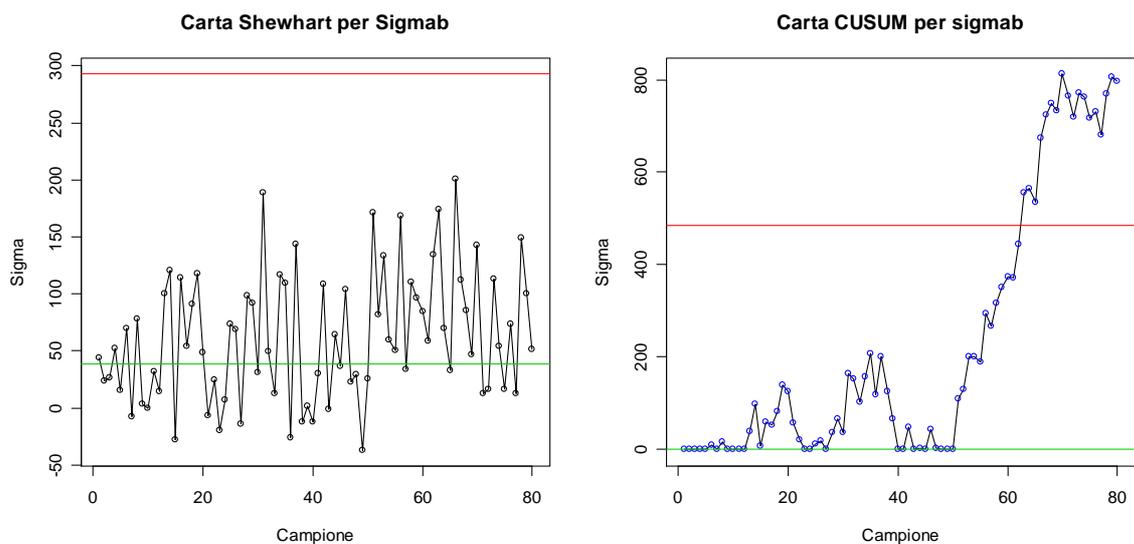


Figura 4.6: Confronto tra la carta Shewhart e CUSUM per σ_b pari a 8.76. La carta Shewhart è costruita con i limiti $UCL = 293.2$, $CL = 38.7$, la carta CUSUM è costruita con una soglia pari a 485.

Analizzando la figura 4.6, come per l'esempio dell'altra componente, la carta Shewhart si rivela meno efficiente per piccoli scostamenti, e non presenta indicazione di cambio della variabilità (la centralità risulta leggermente spostata). La carta CUSUM, implementata per essere sensibile a piccoli cambiamenti in varianza, presenta una deriva crescente a partire dal 50° campione.

4.3. Confronto per la componente μ

Di seguito viene rappresentato il confronto tra le due carte per quanto riguarda il controllo della media del processo. I primi 30 campioni sono simulati partendo con valori dei parametri $\sigma = 7.135$, $\sigma_b = 7.014$, $\mu = 35$, e per gli altri 50 campioni si ottengono un cambiamento in media proporzionale al suo standard error. Per la carta di controllo Shewhart i limiti calcolati al paragrafo 3.1.3. valgono per ogni esempio riportato, mentre per la carta CUSUM, la soglia h sarà cambiata in modo da garantire uno schema sensibile rispetto al rispettivo cambiamento del parametro μ . I valori di tali soglie sono recuperabili al paragrafo 3.2.3.

Valutiamo il comportamento delle carte con uno shift pari a 1.5 volte lo standard error di μ . La nuova media del processo avrà valore pari a 40.79592.

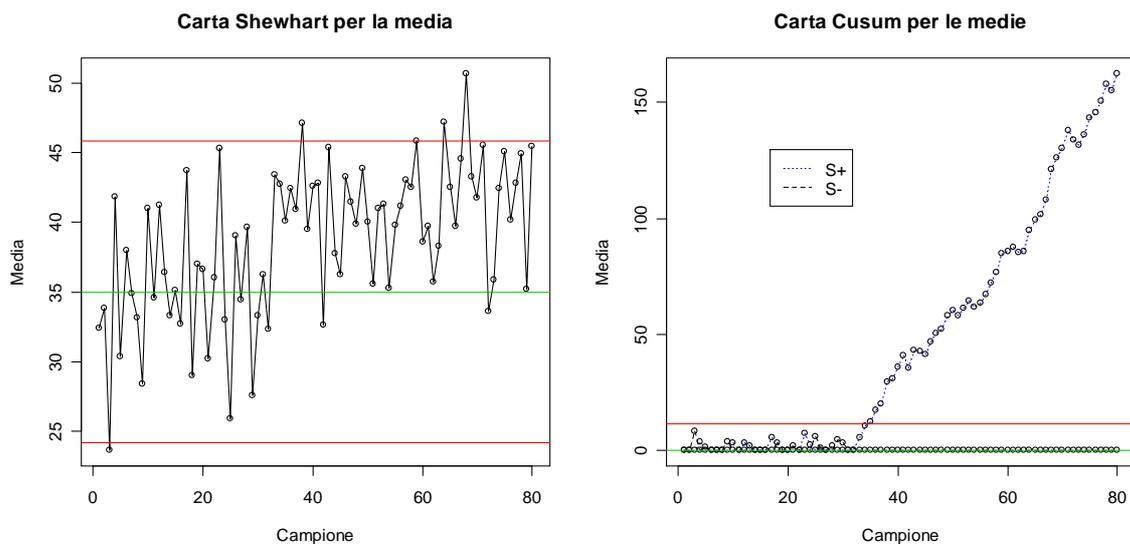


Figura 4.7: Confronto tra la carta Shewhart e CUSUM per μ pari a 40.79. La carta Shewhart è costruita con i limiti $UCL = 45.84$, $CL = 35$, $LCL = 24.15$, la carta CUSUM è costruita con una soglia pari a 11.28.

In questo esempio la carta Shewhart si comporta bene, segnalando fuori controllo subito dopo il 30° campione. Si nota in questo caso che la centralità della statistica è nettamente spostata a un valore più elevato. La carta CUSUM si comporta in maniera soddisfacente.

Valutiamo il comportamento delle carte con uno shift pari a 1 lo standard error di μ . La nuova media del processo avrà valore pari a 38.86394.

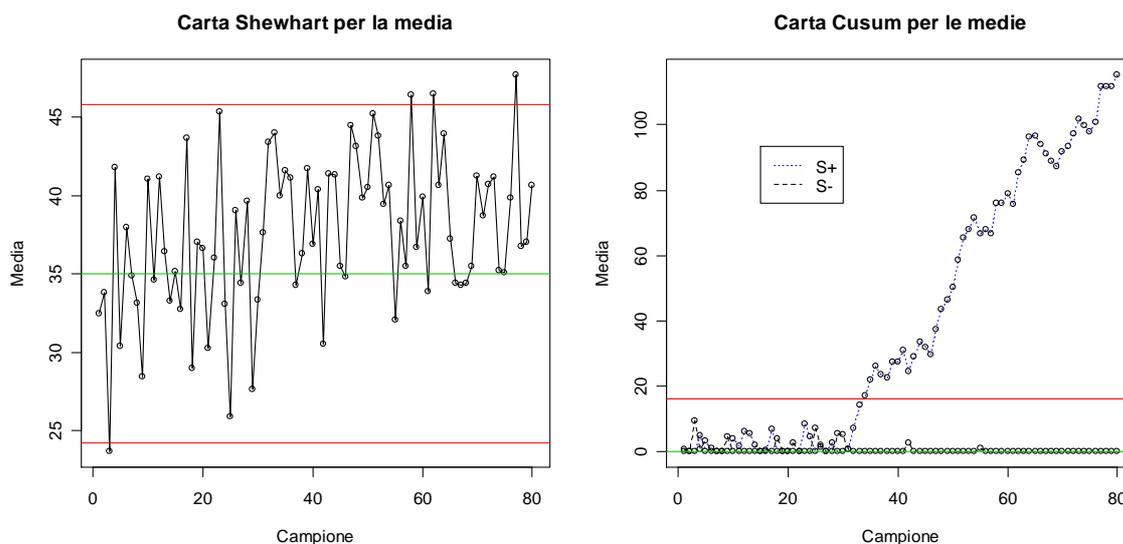


Figura 4.8: Confronto tra la carta Shewhart e CUSUM per μ pari a 38.86. La carta Shewhart è costruita con i limiti $UCL = 45.84$, $CL = 35$, $LCL = 24.15$, la carta CUSUM è costruita con una soglia pari a 16.22.

La carta di controllo Shewhart segnala il primo fuori controllo all'altezza del 58° campione. Si può notare lo spostamento della centralità dei dati appena dopo il cambiamento subito dalla media. La carta CUSUM mostra appena dopo il 30° campione una deriva crescente positiva. Questa carta si comporta meglio rispetto l'altra, perchè la statistica CUSUM supera la soglia H prima che la statistica Shewhart superi il limite di controllo UCL. In definitiva le carte si comportano bene tutte e due, anche se la carta con memoria permette di identificare prima lo scostamento della media dal valore target 35.

L'ultimo esempio affrontato prenderà in considerazione un cambiamento in media pari a 0.5 il suo standard error. Il valore μ sarà pari a 36.93197.

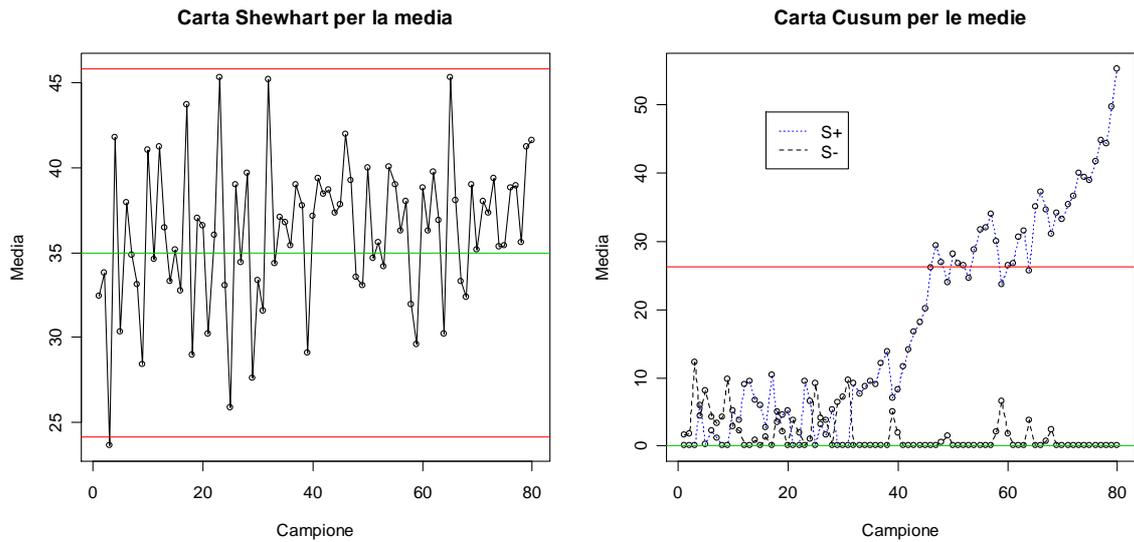


Figura 4.9: Confronto tra la carta Shewhart e CUSUM per μ pari a 36.93. La carta Shewhart è costruita con i limiti $UCL = 45.84$, $CL = 35$, $LCL = 24.15$, la carta CUSUM è costruita con una soglia pari a 26.27.

Con un salto particolarmente piccolo, la carta Shewhart non segnala fuori controllo come si era notato negli altri esempi riguardanti la variabilità. La carta di controllo CUSUM, resa sensibile per scostamenti di piccole dimensioni, si comporta decisamente meglio, presentando una deriva crescente attorno al 30° campione, e segnalando il primo fuori controllo dopo il 40° campione.

CAPITOLO 5

Performance delle carte di controllo

Nell' SPC, è utile predisporre di strumenti in grado di valutare l'efficacia di una carta per individuare scostamenti da valori target. Uno di questi strumenti è l'ARL, definita come il numero medio di campioni occorrente prima che un determinato salto venga individuato. Quando il processo è stabile, l'ARL assume valori elevati in quanto un fuori controllo corrisponderebbe a un falso allarme, mentre quando il processo è instabile, lo schema con un ARL più bassa è migliore, in quanto ci vorranno meno campioni per individuare lo scostamento avvenuto dal valore target. Questo paragrafo affronterà lo studio dell'ARL tra le carte Shewhart, proposte da Chang e Gan, e CUSUM, proposte da Yaschin, in modo da valutarne l'efficacia rispetto a vari cambiamenti in varianza per vari valori di r ed n .

5.1. Performance del sistema di monitoraggio

Una delle più importanti applicazioni delle metodologie descritte, tratta i sistemi automatici per l'SPC volti a rilevare, attraverso le informazioni disponibili, la presenza di condizioni di fuori controllo. Il numero di parametri in questi sistemi di controllo può essere elevato, perciò è importante mantenere il tasso globale di falsi allarmi sotto osservazione. Nel modello descritto al punto 1.3, tre parametri vengono monitorati simultaneamente e, sotto l'assunzione che le ARL target corrispondenti dei tre schemi siano $ARL_{0\mu}$, ARL_{0b} , $ARL_{0\sigma}$, il tasso globale di falsi allarmi può essere approssimato mediante la seguente formula:

$$\frac{1}{ARL_0} \approx \frac{1}{ARL_{0\mu}} + \frac{1}{ARL_{0b}} + \frac{1}{ARL_{0\sigma}}. \quad (5.1)$$

Nelle situazioni pratiche, si può tipicamente fissare il tasso globale di falsi allarmi che può essere tollerato, e poi allocare questo tasso tra gli schemi individuali. Dopo averli disegnati, si dovrebbe simulare l'intero sistema per assicurarsi che il tasso globale di falsi allarmi sia in controllo. Talvolta si rendono necessari lievi aggiustamenti dei livelli di segnale (soglie). La formula (5.1) si può spiegare in questo modo: si supponga che tutti gli schemi vengano fatti partire simultaneamente e, quando uno di essi supera il livello di segnale, il valore S_i per tale schema ritorna a 0 e il monitoraggio viene poi fatto ripartire

(esperimento A). Poichè i parametri sono in target, il risultante tasso di falsi allarmi per i tre schemi sono eventi rari, e sotto condizioni generali hanno una distribuzione approssimata di Poisson. Il combinato tasso di falsi allarmi forma una processo stazionario con tasso equivalente alla parte destra della formula (5.1). Si può considerare un esperimento B, nel quale ogni schema ritorna a 0 quando uno solo di questi supera il livello di segnale. Il combinato tasso di allarmi è equivale alla parte sinistra della (5.1). Il tasso di falsi allarmi nell'esperimento A provvede a una buona approssimazione a quelli dell'esperimento B, poichè esiste una correlazione negativa tra $\hat{\sigma}_{b(k)}^2$ e $\hat{\sigma}_{(k)}^2$. Se nell'esperimento A il valore $\hat{\sigma}_{(k)}^2$ è grande, esso tende a ridurre il valore di $\hat{\sigma}_{b(k)}^2$. In questo modo, nell'istante in cui $\hat{\sigma}_{(k)}^2$ supera la soglia, $\hat{\sigma}_{b(k)}^2$ sarà relativamente piccola rispetto al suo livello di segnale, e l'effetto sarà simile come riportare il valore dello schema a 0. Perciò il tasso di falsi allarmi sia nell'esperimento A, che nell'esperimento B, sono pressappoco equivalenti.

5.2. Performance delle carte di controllo

Questo paragrafo metterà a confronto le carte di controllo Shewhart e CUSUM, per diversi profili ARL. Un profilo ARL più basso risulta migliore in quanto il numero medio di campioni occorrenti per rilevare un cambiamento avvenuto nella specifica componente è minore.

Di seguito si confrontano i profili Shewhart e CUSUM per vari valori di r ed n .

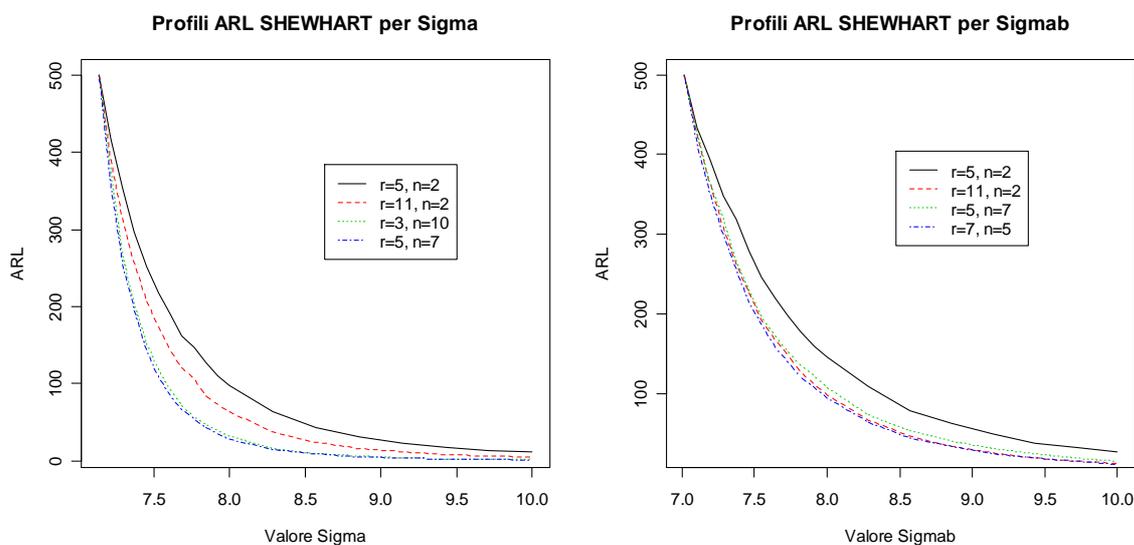


Figura 5.1: Profili della carta Shewhart per le componenti di varianza.

Dalla figura 5.1, si nota che il profilo con $r = 5$ ed $n = 2$, denotato con la linea nera, risulta peggiore rispetto agli altri profili. Quando si sorveglia la componente σ , tendenzialmente il profilo migliora al crescere di n , mentre per σ_b , il profilo migliora al crescere di r , come ci dovremmo aspettare, dato che in un caso controlliamo la componente erratica, e nell'altro la differenza tra le varie postazioni.

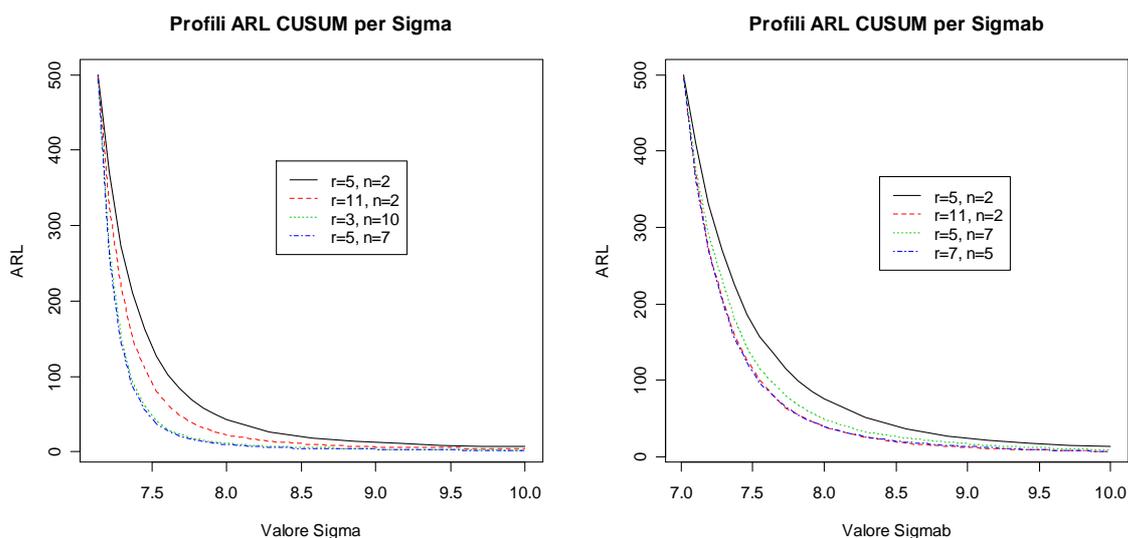


Figura 5.2 Profili della carta CUSUM per le componenti di varianza.

Guardando lo figura 5.2, si possono essere trarre le stesse considerazioni dette per la carta di controllo Shewhart. Per il controllo della componente σ_b , i profili con $r = 11$ e $n = 2$, e $r = 7$ e $n = 5$, sono molto simili. Quindi, un numero equilibrato di r ed n risulta migliore rispetto a un numero r nettamente più grande di n . Questo dipende dal valore assunto da σ_b , per un determinato valore di r ed n , rispetto a un cambiamento proporzionale alla deviazione standard di $\hat{\sigma}_b^2$.

Tabella 1. Per i valori di r ed n elencati, vengono calcolati i salti della componente σ_b pari a una volta il suo standard error, con $\sigma_b = 7.014$.

r	5	11	7	5	3	15
n	2	2	5	7	10	9
Salto di 1 $s.e(\hat{\sigma}_b^2)$	10,21676	9,181549	9,142529	9,442976	10,17354	8,363098

La tabella mostra il salto pari a una volta lo standard error $\hat{\sigma}_b^2$ rispetto a vari valori di r ed n . A parità di salto standardizzato, uno schema che presenta una valore di σ_b minore avrà un ARL più bassa. Per r pari a 15 e n a 9, quando σ_b vale 8.36, per gli altri valori di r ed

n mostrati, il salto standardizzato è minore di 1, quindi l'ARL per questi schemi sarà più alta, poichè lo scostamento da σ_b in termini di deviazioni standard risulta minore. Dalla tabella, senza stilare un profilo ARL, si può dire che per $r = 11$ e $n = 2$ la performance sarà, seppur peggiore, simile a quella per $r = 7$ e $n = 5$ (come si nota dalle figure 5.1 e 5.2).

Il confronto tra la carta Shewhart e CUSUM viene riportato di seguito. Per semplicità, per ogni esempio, è presente un solo profilo ciascuno per le due carte.

Si riportano prima i confronti per la componente σ .

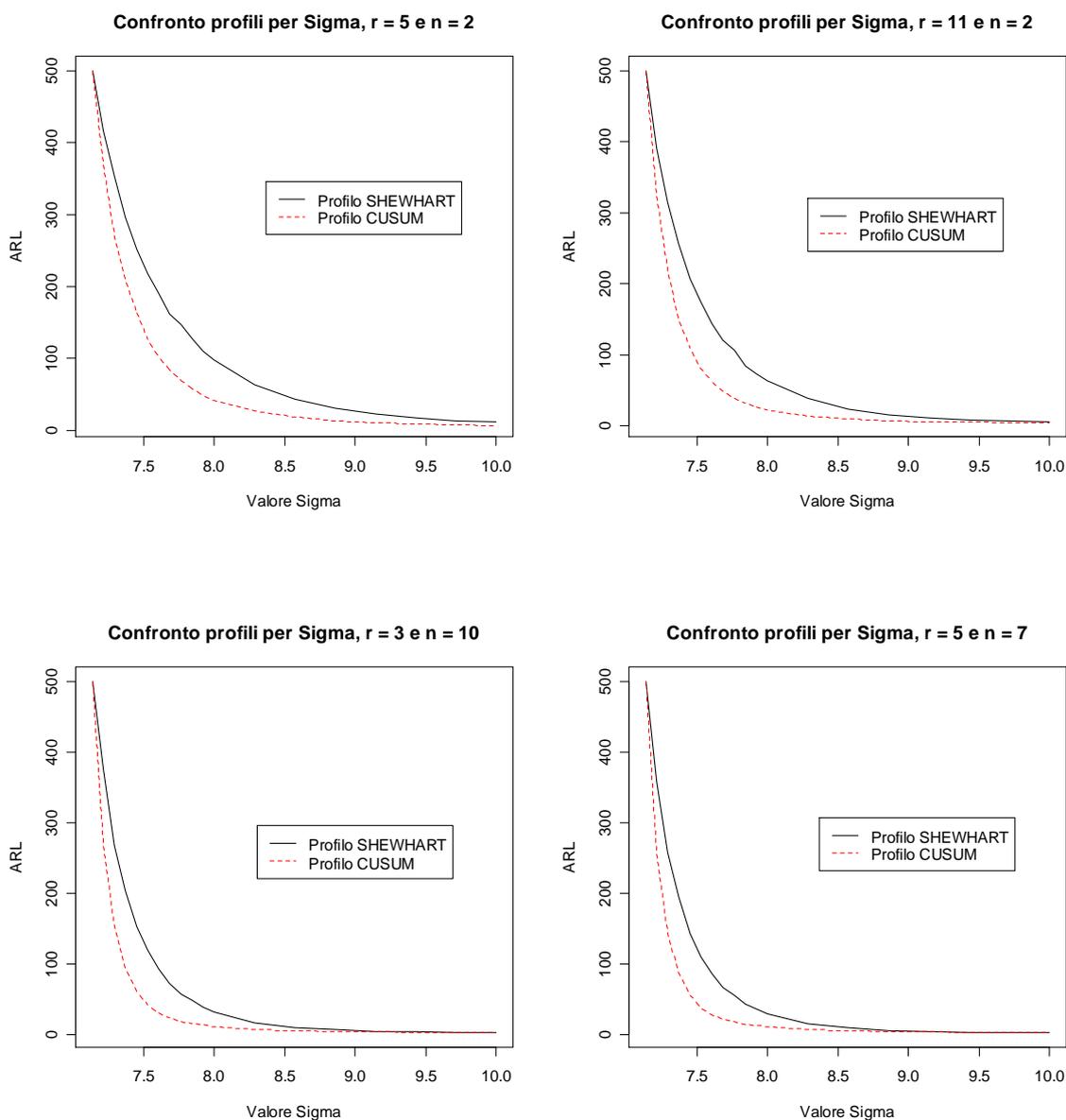


Figura 5.3: Confronti tra la carta Shewhart e CUSUM per la componente σ .

Come si può notare, i profili della carta CUSUM, sensibili per uno scostamento pari a una volta lo standard error di σ , sono più bassi rispetto alla Shewhart. Si può concludere che la carta con memoria risulta più efficiente in quanto, a parità di scostamento della componente, presenta un valore dell' ARL minore.

Di seguito vengono riportati i profili per la componente σ_b .

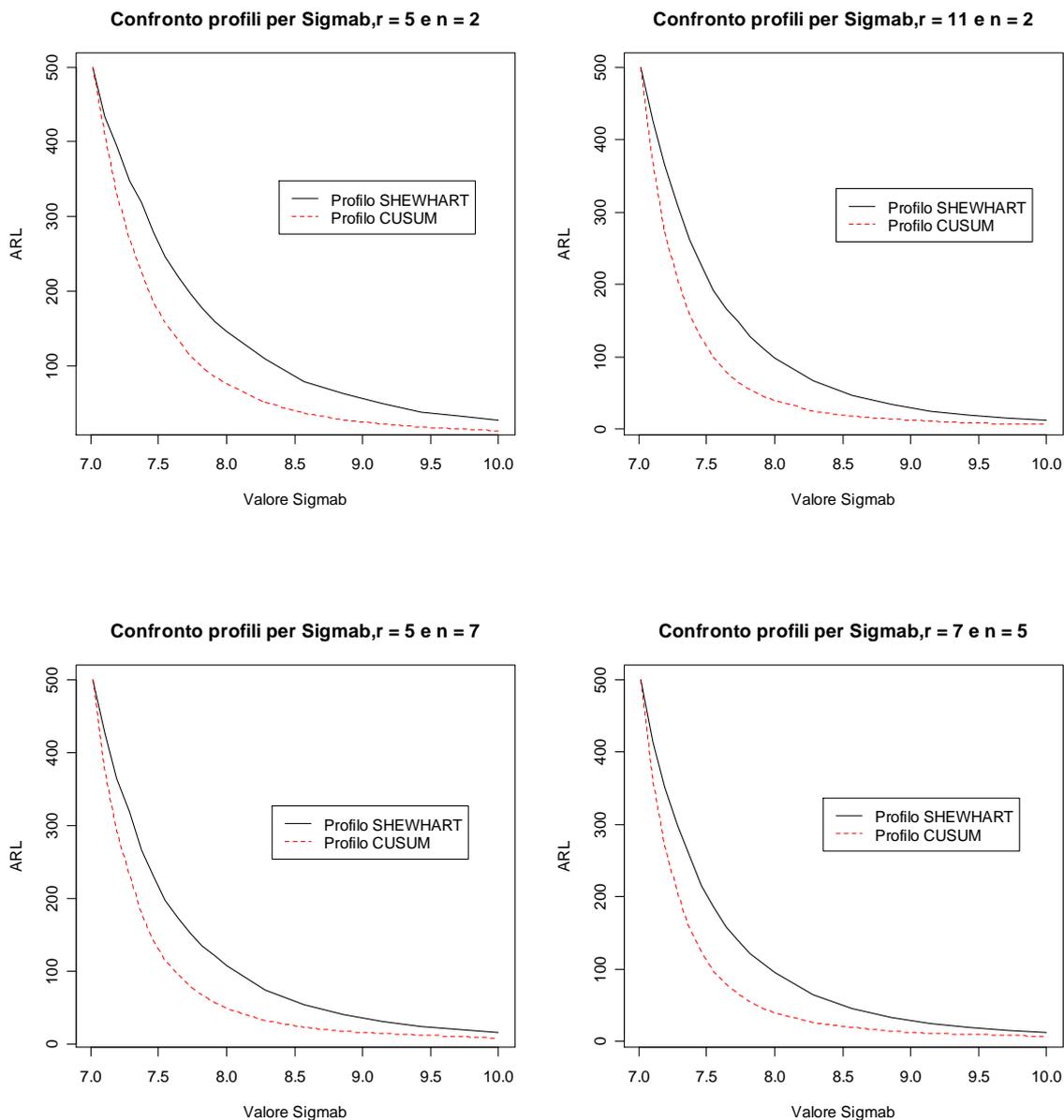


Figura 5.4: Confronti tra la carta Shewhart e CUSUM per la componente σ_b .

Anche per questa componente, lo schema CUSUM risulta più efficace di quello Shewhart. Se si costruiscono carte con memoria sensibili a scostamenti sempre più piccoli, la differenza risulterebbe sempre più evidente.

Abbiamo quindi verificato, tramite i profili ARL, che la carta di controllo CUSUM è più efficace per deviazioni abbastanza piccole delle componenti. La carta di controllo Shewhart, essendo una carta senza memoria, non riesce ad individuare questi scostamenti, mentre la carta con memoria, se costruita per individuare salti piuttosto piccoli, si comporta molto bene, segnalando mediamente prima della carta Shewhart il fuori controllo. Naturalmente, la tecnica CUSUM tende ad essere simile in termine di performance quando viene costruita man mano per salti più grandi. Si guardino i profili sottostanti, per $r = 5$ e $n = 2$, relativi al controllo della componente σ

Confronti profili per Sigma, $r = 5$ e $n = 2$

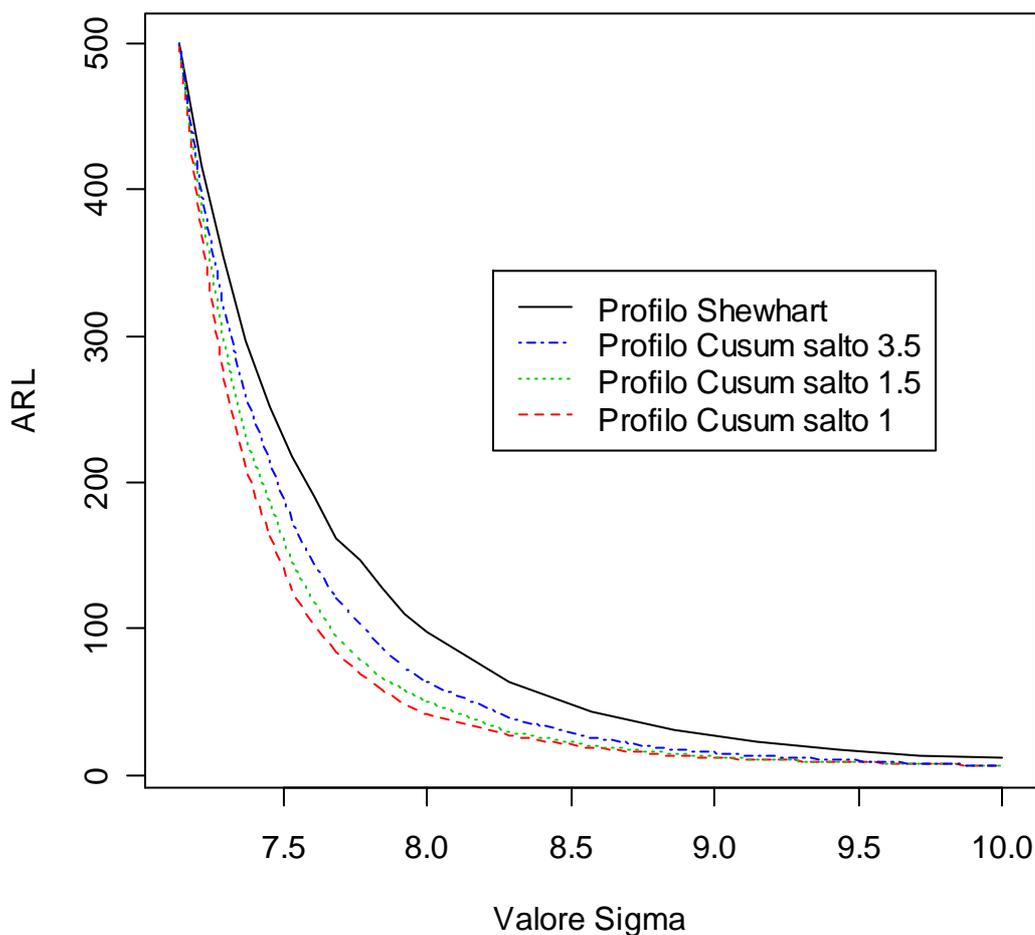


Figura 5.5: Rappresentazione di vari profili per la componente σ , con $r = 5$ e $n = 2$.

In definitiva, conviene utilizzare la carta Shewhart per salti maggiori di 1.5 volte lo standard error della componente ispezionata.

CAPITOLO 6

Carte tradizionali

Le carte di controllo \bar{X} ed S tradizionali sono usualmente utilizzate per controllare la media del processo e la sua variabilità poiché risultano efficienti e facili da interpretare. Tuttavia, in alcune situazioni potrebbero risultare inefficienti se venisse a mancare una delle ipotesi che sta alla base della loro implementazione. Avendo a disposizione campioni di dati, le ipotesi classiche si basano sul concetto che l'intera variabilità del processo risulta strettamente dipendente da quella all'interno dei campioni: in pratica la varianza globale dei dati è spiegata solo dalla differenza delle singole osservazioni rispetto alla media di gruppo. In presenza di ulteriori fattori di variabilità tra i dati, come del resto l'esempio reale affrontato in questo articolo, le tradizionali carte \bar{X} e S si rivelano inadeguate, producendo un numero inaccettabile di falsi allarmi e portando talvolta a confusione, non riuscendo a distinguere se il processo stia lavorando sotto parametri in controllo o meno. Lo scopo di questo capitolo è far osservare l'utilità delle decomposizioni della varianza in più componenti rispetto alle carte tradizionali, valutando i principali vantaggi.

6.1. Carte Shewhart tradizionali

Il controllo della variabilità delle carte Shewhart tradizionali viene fatto tramite le carte S^2 (basata sulla varianza campionaria), oppure la carta R (range campionari). In questo paragrafo verrà studiata la carta sulla varianza campionaria basata sui singoli campioni. Riesamineremo l'esempio portato a inizio articolo, dove per ogni campione abbiamo a disposizione r postazioni con n misure prelevate in ciascuno di queste.

6.1.1. Carta basata su S^2

La carta S^2 tradizionale è volta a monitorare la variabilità all'interno dei campioni. Avendo a disposizione $r \times n$, si potrebbe stimare la varianza all'interno del campione utilizzando la formula seguente:

$$\widehat{S}_k^2 = \frac{\sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \overline{X}_{..})^2}{rn - 1}.$$

Qualora si dovesse stimare la variabilità totale, la si può costruire mediante:

$$\widehat{S}^2 = \frac{\sum_{k=1}^m \widehat{S}_k^2}{m}, \text{ media delle varianze campionarie tra i campioni.}$$

Per la costruzione dei limiti di controllo, si possono determinare i limiti probabilistici basati sulla distribuzione della varianza campionaria σ^2 .

Sfruttando che $\frac{(rn-1)\widehat{S}^2}{\sigma^2} \sim \chi_{m-1}^2$, i limiti di controllo per le sequenze $\{\widehat{\sigma}_k^2\}$ diventano

ora:

$$UCL = \frac{\sigma^2 \chi_{1-\alpha/2, m-1}^2}{rn-1} \quad LCL = \frac{\sigma^2 \chi_{\alpha/2, m-1}^2}{rn-1}, \quad CL = \frac{\sigma^2 \chi_{0.5, m-1}^2}{rn-1},$$

dove al posto di σ^2 può essere sostituita una stima costruita con dati in controllo. La statistica presa sotto esame per ogni campione è l'usuale stima della varianza campionaria \widehat{S}^2 .

6.1.2. Carta per le medie

Le carte Shewhart tradizionali per le medie basate sulla normalità sono semplici disegnare. La statistica presa in esame sarà la media delle medie dei gruppi di un campione (la media delle medie delle r postazioni), vale a dire $\overline{X}_{..}$. I limiti di controllo si ottengono nel modo seguente:

$$UCL = \mu + Z_{1-\alpha/2} \frac{\sigma}{\sqrt{rn}}, \quad CL = \mu, \quad LCL = \mu - Z_{1-\alpha/2} \frac{\sigma}{\sqrt{rn}}.$$

Al posto dei valori μ , σ si possono sostituire delle stime non distorte di tali valori.

6.2. Applicazione

Consideriamo due diversi processi nei quali abbiamo in entrambi $\mu = 35$ e $\sigma = 1$, con σ_b pari a 0 (processo A) e σ_b pari a 1 (processo B), con $r = 7$ e $n = 5$. Lo scopo di questo paragrafo è di valutare il comportamento delle carte Shewhart tradizionali rispetto a quello delle carte la cui varianza viene decomposta.

Consideriamo il primo processo non avente la variabilità tra le postazioni in un campione. Con dei dati simulati, otteniamo le seguenti stime:

$$\hat{\sigma} = 1.008257, \hat{\mu} = 35.02241.$$

Sotto vengono riportate le carte di controllo .

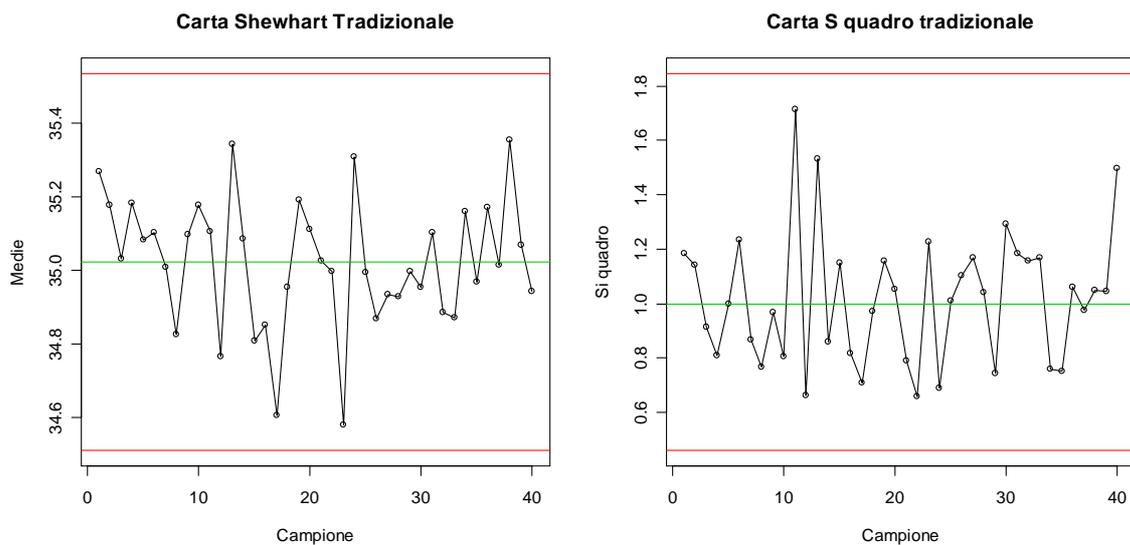


Figura 6.1: carte di controllo Shewhart per la media e S^2 riguardanti il processo A.

Le carte di controllo tradizionali si comportano bene quando la variabilità globale dipende solo dalla differenza delle osservazioni rispetto alle medie delle postazioni cui appartengono.

Valutiamo il comportamento in presenza di un'ulteriore fonte di variabilità. In questo caso, per il processo B, le stime dei parametri risultano: $\hat{\sigma} = 1.365331$, $\hat{\mu} = 35.0395$.

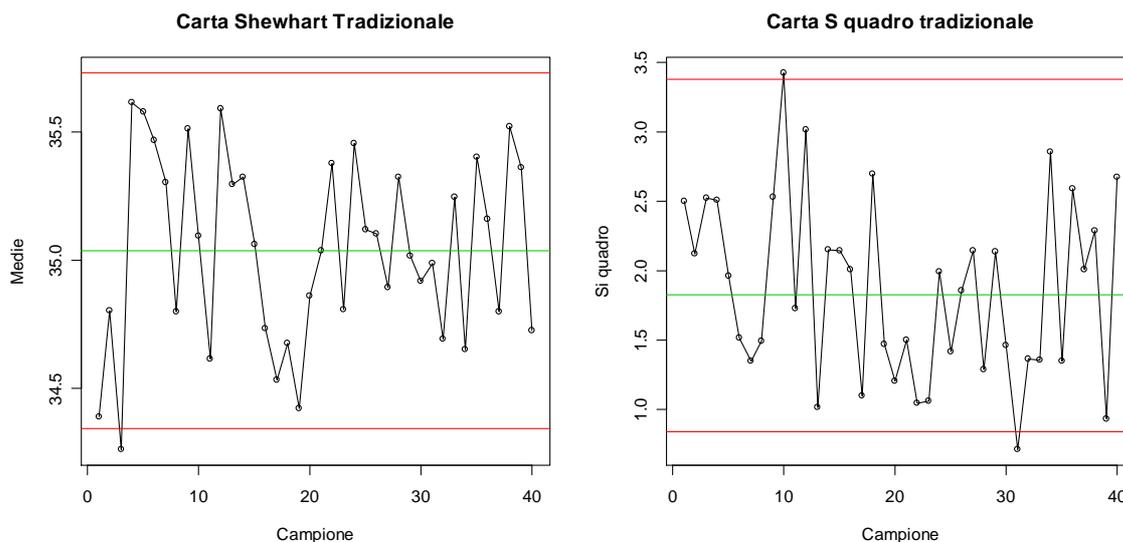


Figura 6.2: Carte di controllo Shewhart per la media e S^2 riguardanti il processo B. I limiti per la media, che dovrebbero mantenere un ARL in controllo pari a 370, sono rispettivamente $UCL = 35.73185$, $CL = 35.0395$, $LCL = 34.34715$, e per la componente σ , che dovrebbero mantenere un ARL in controllo pari a 200, sono rispettivamente $UCL = 3.384936$, $CL = 1.827707$, $LCL = 0.842585$.

Come si può notare dalla figura 6.2, le carte segnalano alcuni fuori controllo che sono da attribuire a falsi allarmi. Quindi risultano inefficienti quando la variabilità totale delle osservazioni dipende da più componenti, in questo caso composta dalla variabilità tra le postazioni e quella entro le postazioni all'interno di un campione. Il procedimento per la costruzione della carta S^2 , come per quello della media, risulta errato.

6.2.1 Valutazioni

Valutiamo il motivo per cui la carta S^2 presenta comportamenti indesiderati. La stima della varianza campionaria usata

$$\widehat{S}_k^2 = \frac{\sum_{i=1}^r \sum_{j=1}^n (X_{ij} - \bar{X}_{..})^2}{rn - 1},$$

viene costruita attraverso lo scarto che intercorre tra la singola osservazione e la media generale del campione. In presenza della sola variabilità σ^2 , la stima tramite \widehat{S}^2 risulta corretta. Tuttavia in presenza di una variabilità tra i gruppi, lo scarto $(X_{ij} - \bar{X}_{..})^2$ diventa più grande, poiché la media generale ha un'ulteriore componente non considerata, l'effetto di σ_b^2 . Infatti sotto il modello ipotizzato abbiamo

$$\bar{X}_{..} = \mu + \frac{1}{r} \sum_{i=1}^r Li + \frac{1}{rn} \sum_{i=1}^r \sum_{j=1}^n \varepsilon_{ij} ,$$

dove la componente Li dipende dalla presenza di σ_b^2 . Non avendo considerato l'esistenza di una variabilità tra le medie delle postazioni, ne risulta che \widehat{S}^2 sovrastima la quantità σ^2 per ogni campione. Come conseguenza la statistica fuoriesce dai limiti di controllo perchè, introducendo nella stima di σ^2 un'ulteriore fonte di variabilità, abbiamo sovrastimato la varianza interna ai gruppi di lotti. In questo caso si parla di falsi allarmi, poiché realmente la componente è in controllo.

Abbiamo visto che anche la carta costruita sulle medie dei campioni presenta una performance non soddisfacente. Il motivo di ciò si può spiegare nella seguente maniera: la variabilità della media di un campione $\bar{X}_{..}$, attorno al suo valor medio μ , misurata come salto in deviazioni standard è pari a

$$\frac{\sigma}{\sqrt{rn}} .$$

Non avendo considerato la presenza di un'ulteriore fonte di variabilità, ne consegue che i limiti di controllo posti a ridosso di μ sono troppo piccoli, in quanto realmente la media $\bar{X}_{..}$ è più variabile rispetto a quanto ipotizzato. La conseguenza è che la statistica più volte supera tali limiti, pur avendo i parametri sotto controllo, provocando così diversi falsi allarmi.

Quando cominciarono a presentarsi i primi problemi con l'utilizzo di queste carte, fu pratica diffusa aumentare i limiti di controllo per una costante arbitraria in modo da ovviare al problema. In questo caso l'effetto indesiderato si presentava quando alcune deviazioni dei parametri dal valore target non venivano rilevate, poiché i limiti venivano alzati troppo.

Riportiamo qui sotto le carte Shewhart studiate in questo articolo, in base alle seguenti stime:

$$\hat{\sigma} = 0.9827054, \hat{\sigma}_b = 1.009062, \hat{\mu} = 35.0395 .$$

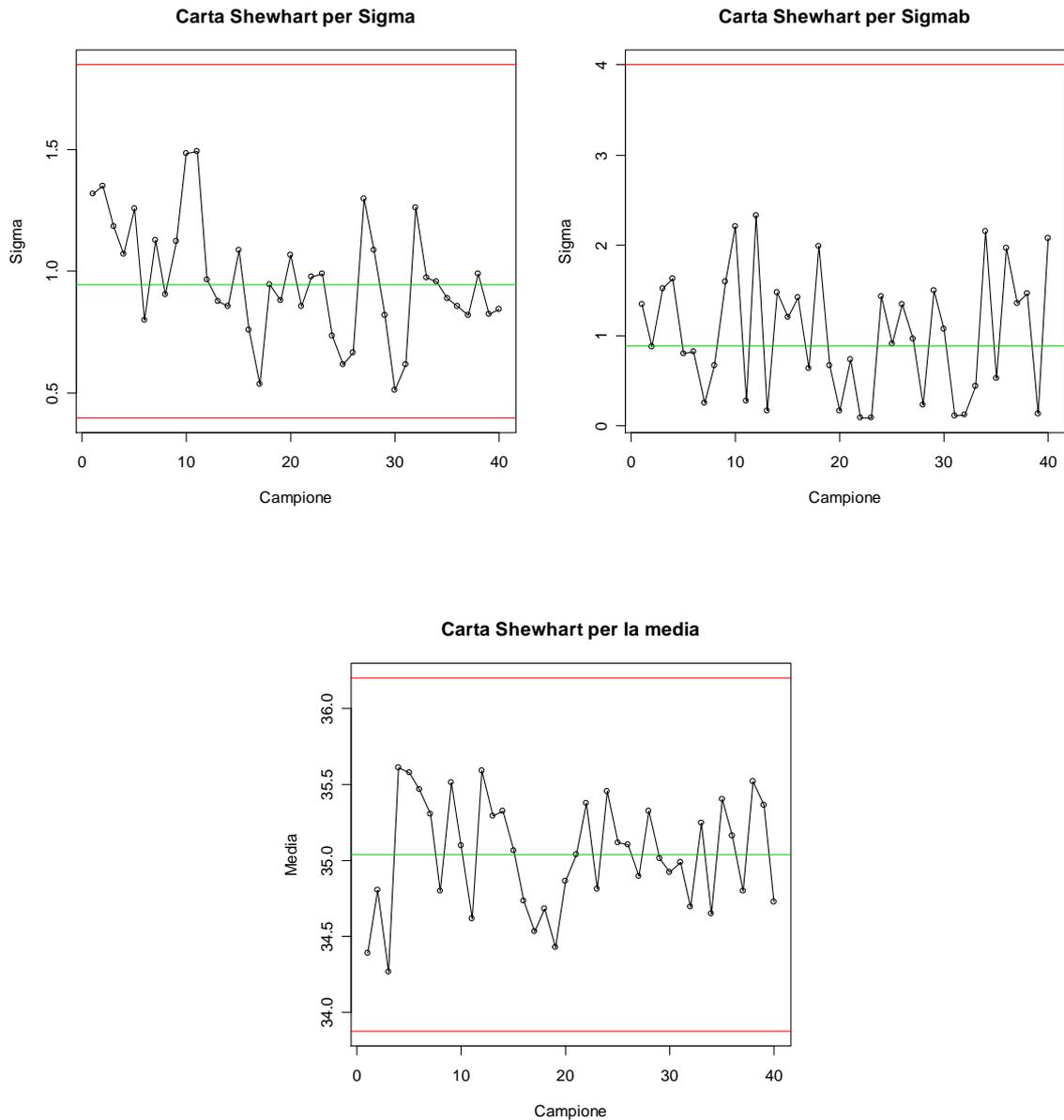


Figura 6.3: Carte di controllo Shewhart per le componenti di varianza. I limiti di controllo per la media sono rispettivamente $UCL = 36.20721$, $CL = 35.0395$, $LCL = 33.87179$, per la componente σ $UCL = 1.84$, $CL = 0.9428167$, $LCL = 0.3965385$, e per la componente σ_b $UCL = 4.008015$, $CL = 0.8877674$.

Come si nota con la decomposizione della variabilità e l'aggiustamento della carta per le medie dei campioni, le carte di controllo non segnalano nulla di rilevante, mostrando così una differenza lampante rispetto alle carte tradizionali.

Concludendo, quando si a motivo di credere che esista una variabilità tra i campioni, quindi non solo all'interno di essi, è utile decomporre la varianza e implementare delle

carte per le singole componenti. In presenza della sola componente di variabilità entro le postazioni, la stima di σ^2 tramite \widehat{S}^2 usando la varianza campionaria può essere fatta .

6.3. Verifica della presenza di ulteriori fonti di variabilità

Naturalmente, qualora fossimo in presenza di ulteriori di fonti di variabilità non considerate nel nostro modello di analisi, degli opportuni accorgimenti possono essere adottati mediante analisi descrittive o test sulla varianza dei campioni.

Riconsideriamo l'esempio apportato al punto 6.2. L'istogramma delle medie standardizzate può riscontrare delle anomalie se l'ipotesi della varianza del modello non è opportuna. Nel processo A, le variabili

$$\frac{\overline{X}_k - \mu}{\sigma_{1c}} \sim N(0,1), \text{ dove } \sigma_{1c} = \sigma / \sqrt{rn} .$$

Con la sola presenza della componente erratica, ci aspettiamo di osservare un andamento delle medie standardizzate simile a quello di una normale. Considerando il processo B, le variabili

$$\frac{\overline{X}_k - \mu}{\sigma_{1c}} \sim N(0, \sigma_{2c} / \sigma_{1c}), \text{ } \sigma_{1c} = \sigma / \sqrt{rn} \text{ e } \sigma_{2c} = \sqrt{(\sigma_b^2 / r) + (\sigma^2 / rn)} .$$

Nella situazione più generale di trovarci con una variabilità globale dipendente da più fattori, ci aspettiamo di osservare un andamento centrale meno denso rispetto a quello della normale, con delle code più pesanti agli estremi. Ovviamente, la differenza sarà tanto più grande quanto il rapporto $\sigma_{2c} / \sigma_{1c}$ sarà più distante dal valore 1.

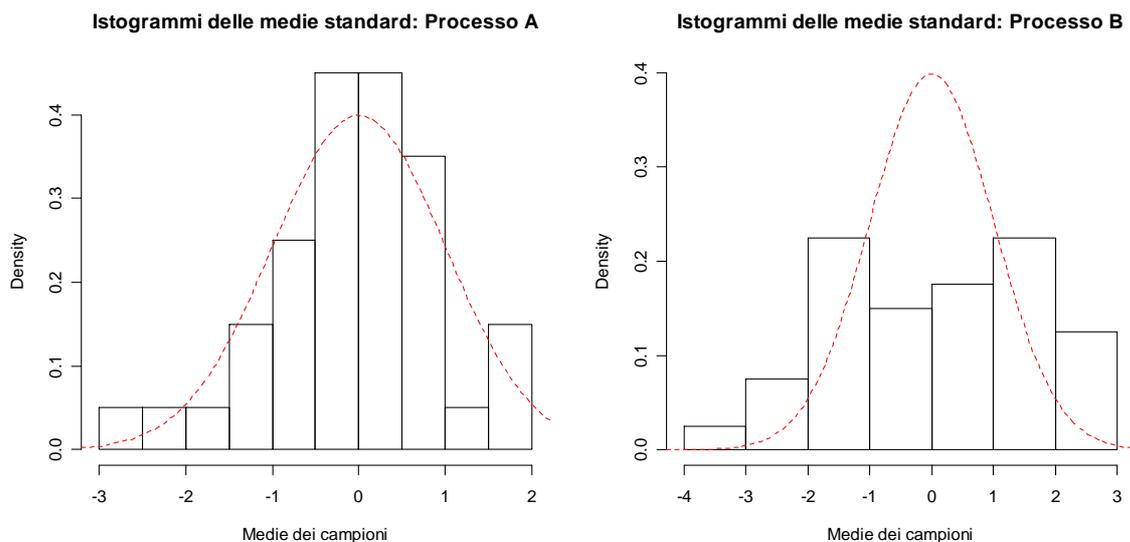


Figura 6.4: Istogrammi delle medie standardizzate per σ_{1c} .

Il grafico a sinistra della figura 6.4 mostra un andamento pressoché normale delle medie standardizzate. Il grafico a destra presenta delle code più pesanti agli estremi, e un addensamento più schiacciato al centro rispetto alla normale standard. Come accennato prima, per valori del rapporto $\sigma_{2c} / \sigma_{1c}$, la differenza risulta più visibile, in quanto la variabilità è molto più alta di quanto ipotizzata. La rappresentazione tramite questi istogrammi risulta abbastanza efficace per valutare la correttezza del modello proposto al punto 1.3.

Un'altra analisi descrittiva può essere fatta attraverso la proiezione dei boxplot delle postazioni per ogni campione.

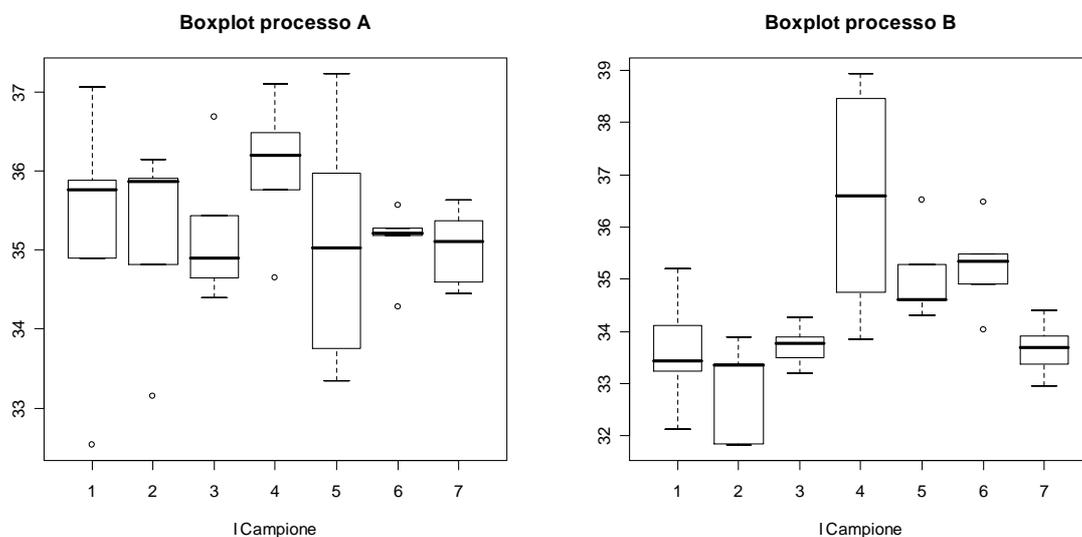


Figura 6.5: Confronto tra il processo A e B tramite i Boxplot delle postazioni relative al primo campione

Per ragioni di spazio, nella figura 6.5 sono stati riportati i boxplot relativi al primo campione per i due processi analizzati. Questa analisi è volta a riscontrare la differenza in media tra le postazioni di un determinato campione. Si possono valutare dei grafici relativi alla distanza delle medie di gruppo rispetto alla media complessiva per ciascun campione.

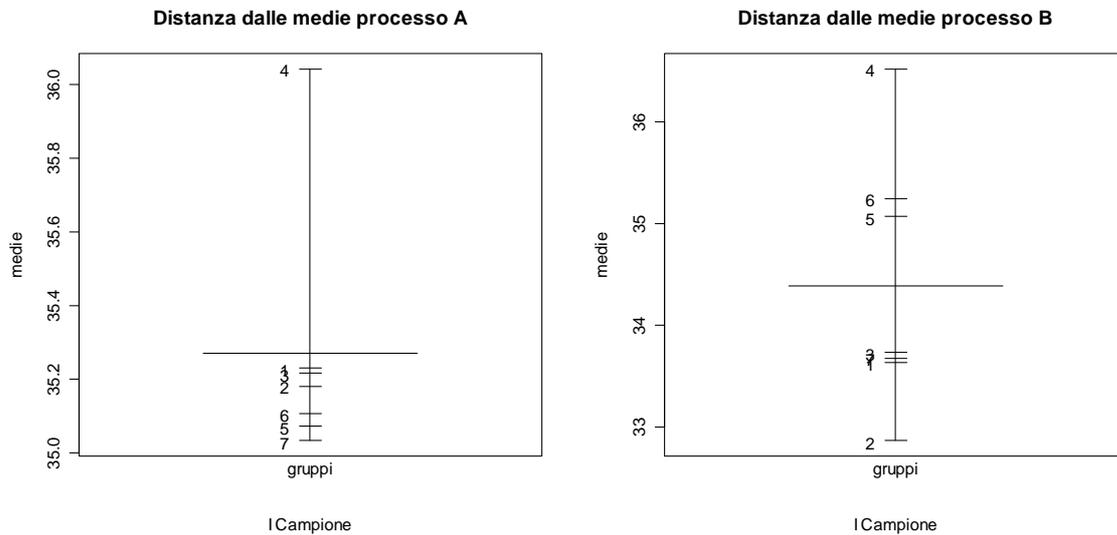


Figura 6.6: Confronto tra il processo A e B mediante la posizione delle medie delle postazioni rispetto alla media complessiva relativa al primo campione.

Questi grafici valutano le differenze delle medie delle postazioni rispetto alla media generale (la linea più grande). Se un processo presenta variabilità in media, la dispersione delle medie di gruppo rispetto a quella totale risulta più elevata, come si nota nel processo B rispetto al processo A.

Anche dei test mediante analisi della varianza possono risultare efficaci per determinare se esista una differenza tra i gruppi, utilizzando la seguente decomposizione:

$$\sum_{k=1}^m \sum_{j=1}^n (X_{kj} - \bar{X}_{..})^2 = n \sum_{k=1}^m (\bar{X}_k - \bar{X}_{..})^2 + \sum_{k=1}^m \sum_{j=1}^n (X_{kj} - \bar{X}_k)^2,$$

dove in questo caso $\bar{X}_{..}$ rappresenta la media complessiva dei dati, \bar{X}_k la media di una singola postazione. Il test statistico mediante analisi della varianza tende a valutare se esiste una differenza significativa tra le medie dei campioni, rispetto alla componente erratica. Il test è stato condotto ipotizzando che un campione fosse costituito da $n = 5$ osservazioni, con $m = 280$ campioni totali a disposizione.

Tabella 2. Confronto dei processi tramite analisi della varianza.

Processo A	Df	Sum Sq	Mean Sq	F value	Pr(>F)	Processo B	Df	Sum Sq	Mean Sq	F value	Pr(>F)
group	279	278,4	1	0,9754	0,5962	group	279	1634,96	5,86	6,6081	< 2,2 e-16
Residuals	1129	1145,78	1,02			Residuals	1129	1081,6	0,97		

Facendo riferimento ai p-value, la tabella sinistra relativa al processo A mostra che non esiste statisticamente una differenza tra le postazioni, come invece avviene per quanto riguarda la tabella a destra riguardante il processo B. Da questa breve analisi si può dedurre che nel processo B è presente una differenza significativa tra le medie delle postazioni, che dovrà essere considerata nel modello di varianza.

6.4. Confronti tra le carte tramite i profili ARL

Come abbiamo visto, le carte Shewhart producono un numero elevato di falsi allarmi quando la variabilità viene considerata erroneamente. Un modo per quantificare la differenza intercorrente tra i due metodi di decomposizione, si basa sullo studio delle ARL. Come esempio prenderemo in considerazione le carte Shewhart basate sulle medie. Questo paragrafo fornirà un confronto tramite ARL tra le carte con la varianza decomposta, e quelle senza.

In generale l'ARL, per le carte di controllo Shewhart basata sulle medie, si calcola tramite il rapporto $1/(1 - \beta)$, dove β esprime l'errore del II tipo definito come:

$$\beta = P(LCL \leq \bar{X}_k \leq UCL | \mu = \mu_0 + \delta),$$

con δ che rappresenta lo scostamento dal valore target μ_0 .

In presenza della sola componente erratica, e δ pari a 0, la generica variabile

$$\frac{\bar{X}_k - \mu}{\sigma_{1c}} \sim N(0,1),$$

e l'ARL in controllo risulta pari a $1/\alpha$.

Quando si verifica uno spostamento dal valore della media di δ unità, la generica variabile

$$\frac{\bar{X}_k - \mu}{\sigma_{1c}} \sim N(\delta / \sigma_{1c}, 1).$$

L'errore del II tipo risulta:

$$\beta = \Phi\left(Z_{1-\alpha/2} - \frac{\delta}{\sigma_{1c}}\right) - \Phi\left(-Z_{1-\alpha/2} - \frac{\delta}{\sigma_{1c}}\right).$$

In presenza di ulteriori fonti di variabilità, la generica variabile

$$\frac{\bar{X}_k - \mu}{\sigma_{1c}} \sim N(0, \sigma_{2c} / \sigma_{1c}).$$

Posto $\vartheta = \sigma_{2c} / \sigma_{1c}$, l'errore del II tipo risulta:

$$\beta = \Phi\left(\frac{Z_{1-\alpha/2}}{\vartheta}\right) - \Phi\left(-\frac{Z_{1-\alpha/2}}{\vartheta}\right).$$

L'ARL apparentemente in controllo diminuisce al crescere del rapporto ϑ . Se si presentasse un cambiamento nella media di δ unità, la generica variabile

$$\frac{\bar{X}_k - \mu}{\sigma_{1c}} \sim N(\delta / \sigma_{1c}, \vartheta),$$

e l'errore del II tipo diventa:

$$\beta = \Phi\left(\frac{Z_{1-\alpha/2}}{\vartheta} - \frac{\delta}{\sigma_{2c}}\right) - \Phi\left(-\frac{Z_{1-\alpha/2}}{\vartheta} - \frac{\delta}{\sigma_{2c}}\right).$$

Quindi le differenze tra le carte aumentano in modo sostanziale all'aumentare della variabilità tra le postazioni.

Di seguito vengono riportate delle tabelle, che con diversi valori di r ed n , presentano i profili ARL per la carte sulla media tra i due tipi di schemi.

Tabella 3. Confronti dell'ARL tra le carte tradizionali e quelle discusse in questo articolo, per vari valori di r ed n . σ_μ rappresenta la stima della varianza delle medie.

Varianza decomposta	Si	No	Si	No	Si	No
σ	1	1	1	1	1	1
σ_b	1	1	0,5	0,5	1	1
σ_μ	0,415704	0,232574	0,256742	0,185369	0,288886	0,157522
r	7	7	7	7	15	15
n	5	5	5	5	5	5
μ	ARL	ARL	ARL	ARL	ARL	ARL
35	200	8,707066	200	24,9086	200	8,475987
35,05025	187,9619	8,509141	171,0116	22,68734	176,4514	8,079351
35,1005	158,6399	7,967215	117,0163	17,84234	128,7621	7,090029
35,15075	124,72	7,206348	73,75474	13,06668	86,30454	5,901117
35,20101	94,554	6,362765	46,20248	9,420415	56,79873	4,7982
35,25126	70,75244	5,539017	29,56718	6,874998	37,77437	3,891746
35,30151	52,94066	4,792762	19,49208	5,137606	25,65403	3,187141
35,35176	39,87669	4,146261	13,26594	3,947033	17,85261	2,6517
35,40201	30,33408	3,600506	9,322455	3,120129	12,74172	2,247369
35,45226	23,33841	3,146281	6,761358	2,536638	9,326863	1,941698
35,50251	18,17268	2,770827	5,057466	2,118621	6,999626	1,709725
35,55276	14,3243	2,461253	3,897993	1,815227	5,383028	1,532985
35,60302	11,43001	2,205976	3,092505	1,592798	4,23962	1,397974
35,65327	9,232218	1,995167	2,522494	1,428652	3,417225	1,294796
35,70352	7,547349	1,820726	2,112625	1,30719	2,816602	1,216126
35,75377	6,24367	1,676077	1,814004	1,217451	2,371926	1,156459
35,80402	5,225883	1,555916	1,594251	1,151542	2,038807	1,111586
35,85427	4,424472	1,455968	1,43148	1,103643	1,78682	1,078232
35,90452	3,788298	1,372787	1,310587	1,069352	1,59478	1,053806
35,95477	3,279427	1,303574	1,22092	1,045277	1,447694	1,036234

Come si può notare dalla tabella precedentemente esposta, quando il rapporto ϑ tende a diventare grande, la differenza tra i due schemi risulta ben visibile. Guardando in particolare l'esempio con r pari a 7 e n pari a 5 con $\sigma_b = 1$, mentre l'ARL in controllo con la decomposizione della varianza è di 200, l'altro schema produce un numero di falsi allarmi una volta ogni 9 campioni. Quindi dalle tabelle emerge che ipotizzando un'unica fonte di variabilità, il numero di questi falsi allarmi risulta troppo elevato.

Tabella 4. Confronti dell'ARL tra le carte tradizionali e quelle discusse in questo articolo, per vari valori di r ed n . σ_μ rappresenta la stima della varianza delle medie.

Varianza decomposta	Si	No	Si	No	Si	No
σ	1	1	1	1	1	1
σ_{ab}	0,5	0,5	0,5	0,5	0,3	0,3
σ_μ	0,166083	0,123453	0,267594	0,153776	0,179912	0,144497
r	15	15	5	5	5	5
n	5	5	10	10	10	10
μ						
35	200	22,01621	200	9,728991	200	26,69644
35,05025	141,4724	18,3276	173,0326	9,175702	148,0735	22,81253
35,1005	70,62903	12,09781	121,2281	7,843411	79,00944	15,75214
35,15075	34,64738	7,621478	78,11141	6,327375	40,64333	10,2355
35,20101	18,11965	4,970797	49,79119	4,998417	21,94081	6,762223
35,25126	10,22424	3,431679	32,29481	3,958915	12,62443	4,659617
35,30151	6,223104	2,515525	21,50937	3,182821	7,752267	3,369248
35,35176	4,075092	1,952924	14,74893	2,611857	5,072393	2,556048
35,40201	2,860689	1,598397	10,41644	2,191978	3,527668	2,029837
35,45226	2,143355	1,37136	7,574888	1,881609	2,599647	1,681873
35,50251	1,704925	1,225371	5,668648	1,650739	2,022486	1,448347
35,55276	1,43084	1,132369	4,36224	1,478121	1,653877	1,290587
35,60302	1,257913	1,074498	3,449006	1,348705	1,414397	1,184286
35,65327	1,149438	1,039813	2,799032	1,251724	1,257816	1,113539
35,70352	1,082854	1,020044	2,329011	1,179345	1,155999	1,067495
35,75377	1,043496	1,009446	1,984479	1,125754	1,090978	1,038474
35,80402	1,021423	1,004147	1,729153	1,086544	1,050711	1,020915
35,85427	1,009825	1,00169	1,538409	1,058306	1,026815	1,010795
35,90452	1,004173	1,000638	1,395216	1,038363	1,013367	1,005271
35,95477	1,001634	1,000223	1,287566	1,0246	1,006251	1,002428

Si nota che per r ed n grandi, la differenza tra le stime delle varianze delle medie risulta minore.

CONCLUSIONI

In molti processi manifatturieri, la varianza del processo può essere decomposta in varie componenti. Spesso, a una rispettiva componente di varianza possono essere associate diverse cause speciali (assegnabili). Nei processi nei quali sono presenti diverse fonti di variabilità, è appropriato considerare l'esistenza di diverse componenti, piuttosto che monitorare la varianza globale. Questo articolo ha sviluppato e discusso le carte di controllo CUSUM e Shewhart per le componenti di varianza. Le carte di controllo CUSUM sono molto efficienti, in termini di ARL, rispetto alla Shewhart, per scostamenti delle componenti minori di una volta le loro deviazioni standard. Le carte tradizionali \bar{X} e S^2 producono un numero elevato di falsi allarmi quando la variabilità dipende da più componenti, quindi si rende necessario utilizzare la regola di scomposizione della varianza in questo caso.

Appendice

Funzioni per il calcolo dell'ARL per le carte Shewhart.

Classe calc_sigma: calcola i valori di σ data una matrice di dati.

```
public class calc_sigma
{
double calcolo_sigma(double [][] matrix,int righe,int colonne)
    {
    double [] vett_medie=calc_medier(matrix,righe,colonne);
    double [] risultato = new double [righe];
    int i,j;
    double cont;
    for(i=0;i<righe;i++)
        {
        cont=0;
        for(j=0;j<colonne;j++)
            {
            cont=cont+(matrix[i][j]-vett_medie[i])*(matrix[i][j]-vett_medie[i]);
            }
        risultato[i]=cont/(righe*(colonne-1));
        }
    cont=0;
    for(i=0;i<righe;i++)
        {
        cont=cont+risultato[i];
        }
    return cont;
    }

double [] calc_medier(double [][] matrix,int righe,int colonne)
    {
    int i,j;
    double cont;
    double [] risultato = new double[righe];
    for(i=0;i<righe;i++)
        {
        cont=0;
        for(j=0;j<colonne;j++)
            {
            cont=matrix[i][j]+cont;
            }
        risultato[i]=cont/colonne;
        }
    return risultato;
    }
}
```

Classe calcolo: calcola i valori di σ_b data una matrice di dati.

```
public class calcolo
{
double calc_varml(double [][] matrix,int righe,int colonne)
    {
    int i;
    double cont=0;
    calc_sigma calcolo = new calc_sigma();
    double [] vettore=calcolo.calc_medier(matrix,righe,colonne);
    double mediat=mediatotale(vettore,righe);
    for(i=0;i<righe;i++)
        {
            cont=cont+(vettore[i]-mediat)*(vettore[i]-mediat);
        }
    return cont/(righe-1);
    }

double calc_siglab(double [][] matrix,int righe,int colonne)
    {
    calc_sigma calc = new calc_sigma();
    double sigma = calc.calcolo_sigma(matrix,righe,colonne);
    double vml = calc_varml(matrix,righe,colonne);
    return vml-(sigma/colonne);
    }

double mediatotale (double [] vettore,int righe)
    {
    int i;
    double cont=0;
    for(i=0;i<righe;i++)
        {
            cont=cont+vettore[i];
        }
    return cont/righe;
    }
}
```

Classe genera: generazione delle componenti casuali L_i ed ε_i .

```
import java.util.Random;
public class Genera
{
Random r1;
Random r2;
Genera()
    {
    r1=new Random();
    r2=new Random();
    }

void gen_li(double [] array,double sigma)
```

```

    {
    int i;
    for(i=0;i<array.length;i++)
        array[i]=r1.nextGaussian()*sigma;
    }
void gen_ei(double [] array,double sigma)
    {
    int i;
    for(i=0;i<array.length;i++)
        array[i]=r2.nextGaussian()*sigma;
    }
}

```

Classe simul_valori: simulazione di un campione di dati.

```

public class simul_valori
{
Genera sim = new Genera();
double [][]calcolo (double mu,double sigmab,double sigma,int righe,int colonne)
    {
    double [] vettore1=new double[righe];
    double [] vettore2=new double[colonne];
    double [][] matrice = new double[righe][colonne];
    sim.gen_li(vettore1,sigmab);
    int i,j;
    for(i=0;i<righe;i++)
        {
        sim.gen_ei(vettore2,sigma);
        for(j=0;j<colonne;j++)
            {
            matrice[i][j]=mu+vettore1[i]+vettore2[j];
            }
        }
    return matrice;
    }
}

```

Classe simul_nvolte: simula dati da nvolte campioni e calcola le stime per σ

```

public class simul_nvolte
{
simul_valori sim=new simul_valori();
calc_sigma calc=new calc_sigma();
double [] calcolo(double mu,double sigmab,double sigma,int righe,int colonne,int nvolte)
    {
    double [] valori=new double[nvolte];
    int i;
    for(i=0;i<nvolte;i++)
        {
        double [][] matrice=sim.calcolo(mu,sigmab,sigma,righe,colonne);

```

```

        valori[i]=calc.calcolo_sigma(matrice,righe,colonne);
    }
    return valori;
}
}

```

Classe simul_nvolte_b: simula dati da nvolte campioni e calcola le stime per σ_b .

```

public class simul_nvolte_b
{
    simul_valori sim=new simul_valori();
    calcolo calc=new calcolo();
    double [] calcolo(double mu,double sigmab,double sigma,int righe,int colonne,int
    nvolte)
    {
        double [] valori=new double[nvolte];
        int i;
        for(i=0;i<nvolte;i++)
        {
            double [][] matrice=sim.calcolo(mu,sigmab,sigma,righe,colonne);
            valori[i]=calc.calc_sigmab(matrice,righe,colonne);
        }
        return valori;
    }
}

```

Classe simul_nvolte_medie: simula dati da nvolte campioni e calcola le stime per μ .

```

public class simul_nvolte_medie
{
    simul_valori sim=new simul_valori();
    calc_sigma calc=new calc_sigma();
    double [] calcolo(double mu,double sigmab,double sigma,int righe,int colonne,int nvolte)
    {
        double [] valori=new double[nvolte];
        int i;
        double [] p=new double[righe];
        for(i=0;i<nvolte;i++)
        {
            double [][] matrice=sim.calcolo(mu,sigmab,sigma,righe,colonne);
            p=calc.calc_medier(matrice,righe,colonne);
            valori[i]=mediatotale(p,righe);
        }
        return valori;
    }
}

```

```

double mediatotale (double [] vettore,int righe)

```

```

    {
    int i;
    double cont=0;
    for(i=0;i<righe;i++)
        {
            cont=cont+vettore[i];
        }
    return cont/righe;
    }
}

```

Classe simul_ARL_sh: calcolo dell'ARL per la carta di controllo Shewhart mediante simulazione del modello.

```

public class simul_ARL_sh
{
    calc_ei_sh esh=new calc_ei_sh();
    simul_nvolte_b lsh = new simul_nvolte_b();
    simul_nvolte_medie msh=new simul_nvolte_medie();
    double [] calcolo(double mu,double sigmab,double sigma,int righe,int colonne,int
    nvolte,int qsim,double UCL,double LCL,int scelta)
    {
        int i,j;
        double [] vettore = new double[nvolte];
        double [] conteggio=new double[qsim];
        double s=0;
        for(i=0;i<qsim;i++)
            {
                switch(scelta)
                {
                    case 1:
                        vettore=esh.calcolo(mu,sigmab,sigma,righe,colonne,nvolte);
                        for(j=0;j<nvolte;j++)
                            {
                                if(vettore[j]>UCL)
                                    {
                                        conteggio[i]=j+1;
                                        break;
                                    }
                                else    conteggio[i]=nvolte;
                            }break;
                    case 2:
                        vettore=lsh.calcolo(mu,sigmab,sigma,righe,colonne,nvolte);
                        for(j=0;j<nvolte;j++)
                            {
                                if(vettore[j]>UCL)
                                    {
                                        conteggio[i]=j+1;
                                        break;
                                    }
                            }
                }
            }
    }
}

```

```

        }
        else    conteggio[i]=nvolte;
        }break;
case 3:
vettore=msh.calcolo(mu,sigmab,sigma,righe,colonne,nvolte);
for(j=0;j<nvolte;j++)
    {
        if(vettore[j]>UCL | vettore[j]<LCL)
            {
                conteggio[i]=j+1;
                break;
            }
        else    conteggio[i]=nvolte;
        }break;
    }
}
for(i=0;i<qsim;i++)
    {
        s=s+conteggio[i];
        double [] res={s/qsim,Max(conteggio)};
return res;
    }

double Max(double []vett)
    {
    int i;
    double max=0;
    for(i=0;i<vett.length;i++)
        {
            if(vett[i]>max)
                {
                    max=vett[i];
                }
            else {}
        }
    return max;
    }
}

```

Bibliografia

- Chang, T.C. e Gan, F.F. (1995). “A Cumulative Sum Control Chart for Monitoring Process Variance”, *Journal of Quality Technology*.
- Chang, T.C. e Gan, F.F. (2004). “Shewhart Chart for Monitoring the Variance Components”, *Journal of Quality Technology*.
- Duncan, A. J. (1986). “Quality Control and Industrial Statistics”, Irwin e Homewood.
- Ewan, W.D. (1963). “When and how to use Cusum Charts”, *Technometrics*.
- Gan, F.F. (1991). “An Optimal Design of Cusum Quality control Charts”, *Journal of Quality Technology*.
- Grant, E. L. e Leavenworth, R. S. (1998). “Statistical Quality Control”, McGraw-Hill.
- Hawkins, D.M. (1981). “A Cusum for a Scale Parameter”, *Journal of Quality Technology*.
- Hawkins, D.M. (1992). “A Fast Accurate Approximation of Average Run Lengths of Cusum Control Charts”, *Journal of Quality Technology*.
- Hawkins, D.M. (1993a). “Cumulative Sum Control Charting: An underutilized SPC tool”, *Quality Engineering*.
- Howell, D. (1987). “A Cusum Scheme for the Control of Process Variance”, presented at the 1987 Joint Statistical Meetings, San Francisco.
- Lucas, J.M. (1976). “The design and use of Cumulative Sum Quality Control Schemes”, *Journal of Quality Technology*.
- Montgomery, D.C. (2000). “Il controllo statistico della qualità”, McGraw-Hill.
- Montgomery, D.C. (2001). “Introduction to Statistical Quality Control”, 4 ed. John Wiley & Sons.
- Page, E.S. (1961). “Cumulative Sum of Control Charts”, *Technometrics*.
- Woodall, W.H. e Adams, B.M. (1993). “The Statistical Design of Cusum Charts”, *Quality Engineering*.
- Yaschin, E.(1985). “On a Unified Approach to the Analysis of Two-Sided Cumulative Sum Schemes with Headstarts”, *Advances in Applied Probability*.
- Yashchin, E. (1994). “Monitoring Variance Components”, *Technometrics*.

Ringraziamenti

A conclusione del mio corso di laurea triennale, desidero fare alcuni ringraziamenti a quanti mi hanno sempre sostenuto e incoraggiato. Un ringraziamento particolare va innanzitutto alla mia famiglia, che, tra le altre cose, mi ha permesso di intraprendere questa bellissima esperienza universitaria, e per questo non potrò mai ringraziarli abbastanza. Desidero ringraziare la professoressa Capizzi per la sua disponibilità e pazienza. Ringrazio i miei compagni universitari, in particolare Nicola e Giuseppe per il loro costante sostegno morale. Ringrazio i miei compagni di appartamento, per la pazienza dimostrata nei miei confronti: Max, Chiara, Veronica, Daniela.

Ultimi, ma non per ordine di importanza, ringrazio quegli amici che mi hanno sempre sostenuto: Mezzo, René, Ciano, Francesca, Giuky, Mary, Lorena, Stefanet,.....e molti altri.