

UNIVERSITA' DEGLI STUDI DI PADOVA

FACOLTA' DI SCIENZE STATISTICHE

CORSO DI LAUREA IN STATISTICA E TECNOLOGIE INFORMATICHE



TESI DI LAUREA TRIENNALE

**INDICI DI CAPACITA' DI PROCESSO IN PRESENZA
DI COMPONENTI GERARCHICHE DI VARIAZIONE**

Relatore: Ch.ma Prof.ssa Giovanna Capizzi

Laureando: Daniele Gomiero

ANNO ACCADEMICO 2007/2008

INDICE

INTRODUZIONE	pag. 1
CAPITOLO 1	
1.1 Introduzione sugli indici di capacità e relativa terminologia	pag. 3
1.2 Indici di capacità	pag. 5
1.3 Stima della deviazione standard short-term	pag. 6
1.4 Lacune degli indici di capacità	pag. 7
CAPITOLO 2	
2.1 Introduzione	pag. 9
2.2 Analisi della varianza	pag. 9
2.2.1 ANOVA con effetti fissi	pag. 9
2.2.2 ANOVA annidata	pag. 10
CAPITOLO 3	
3.1 Indici di capacità in presenza di componenti gerarchiche di variazione	pag. 14
3.2 Ancora su indici di capacità in presenza di componenti gerarchiche di variazione	pag. 17
3.2.1 Il modello process set-up	pag. 17
3.2.2 Il modello special cause	pag. 20
CAPITOLO 4	
4.1 Simulazione	pag. 22
CONCLUSIONI	pag. 27
APPENDICE	pag. 28
BIBLIOGRAFIA	pag. 34
RINGRAZIAMENTI	pag. 35

INTRODUZIONE

Già nel lontano 1920 Walter Shewhart iniziò a ideare un nuovo concetto e un nuovo modo di lavorare per le aziende, introducendo quello che noi oggi conosciamo come controllo statistico di processo (SPC).

Nonostante l'SPC sia nato quasi 100 anni fa, possiamo affermare che questo è rimasto un po' in sordina fino a pochi anni fa.

E' infatti con l'inizio degli anni '80 che all'interno delle aziende inizia a svilupparsi l'idea del controllo statistico della qualità. Tale processo ha avuto uno sviluppo esponenziale con il passare del tempo sino ad arrivare al suo apice storico, nei giorni nostri.

Le aziende infatti si sono rese conto dell'importanza di soddisfare i requisiti del cliente ed hanno quindi avviato degli studi sulla capacità dei loro processi in modo da fornire la sicurezza al cliente di quanto promesso, ossia la qualità del loro prodotto.

In questa trattazione mi soffermerò per l'appunto sugli indici di capacità di processo.

Letteralmente la capacità è l'abilità nel compiere un incarico od ottenere un risultato che viene predefinito dalle specifiche richieste del cliente.

Gli indici di capacità sono quindi dei valori che forniscono indicazione di quanto il processo riesca a realizzare dei prodotti le cui caratteristiche appartengono all'intervallo di specifica, stabilito a priori dagli ingegneri del processo, dal progettista o dal cliente stesso.

Il motivo principale per la stesura di questa tesi è stato il mio grande interesse sull'argomento sviluppato seguendo vari corsi all'interno della facoltà di scienze statistiche, in secondo luogo mi ha molto affascinato tale argomento in quanto lo studio degli indici di capacità in condizioni operative sia ancora in fase di sviluppo. Al giorno d'oggi le aziende sono in seria difficoltà nell'adottare gli indici standard alla loro realtà.

La mia tesi è costituita in tre capitoli principali:

Nel primo capitolo tratterò gli indici di capacità in senso generale, fornendo alcuni interessanti cenni storici, informazioni sul loro impiego nel controllo statistico della qualità, nonché sulla loro incapacità di fornire risultati validi in certe situazioni che si vengono a creare nella realtà produttive.

Nel secondo capitolo si descrive brevemente la "nested" ANOVA a una e a due vie, ad effetti fissi e casuali. L'introduzione del metodo ANOVA per la stima della

varianza è di fondamentale importanza per fornire valori attendibili delle varie componenti di variazione presenti nella nostra realtà produttiva.

Il terzo capitolo è sicuramente il capitolo centrale della mia tesi, in cui tratterò gli indici di capacità in presenza di componenti gerarchiche di variazione. Introdurrò delle stime della percentuale di variazione dovuta alle cause speciali e al set-up del processo. Tratterò inoltre lo studio GR&R e un metodo efficace per l'identificazione delle cause che influiscono nel processo portando una elevata varianza.

Nel quarto capitolo effettuo invece uno studio di simulazione in cui metto a confronto l'indice di Kane (1986) e un indice della famiglia di Vannman (1995) in presenza di vari shift della media del processo, tale confronto viene effettuato attraverso grafici (boxplot) e in termini dei quantili della distribuzione di frequenza dei due indici.

Seguono le conclusioni tratte al termine dalla realizzazione di questa tesi, l'appendice in cui ho inserito le funzioni usate per la simulazione dei dati e il calcolo degli indici di capacità relativi alla parte di simulazione, la bibliografia sulla quale mi sono basato e infine i dovuti ringraziamenti.

CAPITOLO 1

1.1 Introduzione sugli indici di capacità e relativa terminologia

Gli indici di capacità come ricordato in precedenza forniscono dei valori per valutare quanto il processo da noi controllato riesca a rispettare i limiti di specifica, inoltre sono di fondamentale importanza per identificare cosa bisogna migliorare all'interno del nostro processo.

In altre parole l'analisi della capacità di un processo è utilizzata per effettuare previsioni su come un processo ben regolato possa rispettare determinate tolleranze; indicare la frequenza di campionamento per il controllo di processo; pianificare le sequenze del processo produttivo quando vi è un effetto sulle tolleranze e per ultima ma non meno importante, ridurre la variabilità del processo di fabbricazione.

L'ultimo punto che noi abbiamo elencato merita una spiegazione più approfondita, infatti in tutti i processi sono presenti due tipi di variazione, la variazione per cause comuni e la variazione per cause speciali.

La variazione per cause comuni è un disturbo che è prevedibile, quasi di routine e per questo non ha alcuna causa assegnabile. Per questi motivi tale variazione è presente in tutti i processi e non può essere eliminata a meno di cambiamenti fondamentali nel processo.

Di ben altra natura è invece la variazione per cause speciali, tale variazione è il risultato della presenza di cause assegnabili, queste devono essere individuate ed eliminate per riportare il processo verso la stabilità, infatti la variazione dovuta a tali cause può indicare un cambiamento qualitativo serio del processo.

Per ottenere le informazioni su come si comporta il nostro processo dobbiamo quindi calcolare gli indici di capacità, ma prima dobbiamo essere a conoscenza degli assunti su cui si basano tali indici, sulla terminologia utilizzata in questi casi e sulle informazioni che possiamo ricavarne. Gli assunti necessari per il calcolo degli indici di capacità tradizionali al fine di evitare la formulazione di conclusioni errate, sono che le osservazioni ricavate dal campione abbiano distribuzione normale, che i dati siano tra loro incorrelati e che il processo sia statisticamente in controllo.

Per poter verificare la validità di tali assunti bisogna effettuare un'analisi preliminare dei dati a nostra disposizione.

La prima operazione da eseguire è lo studio della variabilità di questi dati e della loro tendenza centrale attraverso l'utilizzo di istogrammi e carte di probabilità, verificheremo quindi con l'utilizzo di diagrammi, quali il Normal Probability Plot, e di test, quale il test di Shapiro-Wilk, la distribuzione normale di questi dati, successivamente con l'utilizzo di carte di controllo verificheremo che il processo sia statisticamente in controllo ovvero stabile.

Ricordiamo che un processo risulta stabile se l'output non presenta segni di cause speciali, mentre è instabile se influenzato da cause speciali.

Il processo che risulta ideale per una realtà aziendale è un processo stabile, che produce output all'interno dei limiti di specifica. Tuttavia nella realtà tali processi sono rari e di difficile ottenimento.

E' invece più facile incontrare casi in cui il processo è definito promettente o ingannevole.

Un processo è definito promettente quando risulta stabile ma produce una quantità significativa di prodotti al di fuori dei limiti di specifica. In questi casi bisogna trovare delle soluzioni per ovviare a tali problemi, infatti se le medie del processo risultano lontane dal valore target dobbiamo cercare di avvicinarle al target prestabilito (spesso lo si può fare facilmente), ben più ostico da correggere è un processo che risulta centrato ma la sua variazione è troppo grande.

Un processo invece è definito ingannevole se risulta instabile ma produce output all'interno dei limiti di specifica, in questo caso dobbiamo adottare un sistema di controllo del processo per stabilizzarlo e dobbiamo ridurre la variabilità intorno al target.

Un processo è infine definito turbolento quando risulta instabile e spesso produce degli output che stanno al di fuori dei limiti di specifica, nel caso in cui ci si trovi di fronte a un caso del genere bisogna innanzitutto cercare di stabilizzare il processo e successivamente migliorarlo; bisogna comunque tenere ben presente che un processo instabile è molto difficile da tenere vicino al target.

1.2 Indici di capacità

Gli indici di capacità più noti sono sicuramente l'indice C_p (Juran, 1974) e l'indice C_{pk} (Kane, 1986). Tuttavia esistono altri due indici molto usati ossia l'indice C_{pm} (Hsiang e Taguchi, 1985) e l'indice C_{pmk} (Choi e Owen 1990, Pearn 1992), altro indice di fondamentale importanza è la famiglia $C_p(\mu, \nu)$ proposta da Vanman (1995).

Siano LSL e USL i limiti di specifica e μ e σ_{ST} rispettivamente la media e la variazione di breve periodo della caratteristica di qualità.

L'indice che per primo fu formulato è l'indice C_p , introdotto da Juran (1974) ed inizialmente noto come Capability Ratio, è una misura dell'abilità del processo di fabbricare prodotti conformi alle specifiche stabilite.

E' dato da:

$$C_p = \frac{USL - LSL}{6\sigma_{ST}} \quad (1.1)$$

Il processo è definito capace se l'indice (1.1) risulta maggiore di 1,5; risulta marginalmente capace se è compreso tra 1 e 1,5; risulta non capace se è minore di 1.

Kane (1986) introduce un nuovo indice di capacità, C_{pk} , tale indice tiene conto della posizione della media μ rispetto ai limiti di specifica ed è definito come segue:

$$C_{pk} = \frac{\min(USL - \mu; \mu - LSL)}{3\sigma_{ST}} \quad (1.2)$$

Analogamente con quanto visto per l'indice (1.1) il processo è definito capace se l'indice (1.2) risulta maggiore di 1,5; risulta marginalmente capace se è compreso tra 1 e 1,5; risulta non capace se è minore di 1. Nel caso in cui è minore di 0 significa che la media del processo è al di fuori dei limiti di specifica.

Confrontando i valori degli indici (1.1) e (1.2) il processo risulta centrato sul livello nominale se $C_p = C_{pk}$, mentre risulta fuori centro se $C_{pk} < C_p$.

Hsiang e Taguchi (1985) introducono un nuovo indice, C_{pm} , noto anche come indice di Taguchi, della forma:

$$C_{pm} = \frac{USL - LSL}{6 * \sqrt{\sigma_{ST}^2 + (\mu - T)^2}} \quad (1.3)$$

Tutti gli indici di capacità descritti in precedenza possono essere ricavati come casi particolari della famiglia introdotta da Vanmann (1995):

$$C_p(u, v) = \frac{d - u | \mu - M |}{3 * \sqrt{\sigma_{ST}^2 + v(\mu - T)^2}} \quad (1.4)$$

dove $d=(USL-LSL)/2$, $M=(USL+LSL)/2$ e u e v sono delle costanti positive. In particolare:

$$C_p = C_p(0,0); C_{pk} = C_p(1,0); C_{pm} = C_p(0,1); C_{pmk} = C_p(1,1).$$

Vanmann in particolare suggerisce di scegliere come valore dei parametri $u=0$ e $v=4$. Tale combinazione fornisce infatti risultati soddisfacenti in un'ampia gamma di situazioni pratiche.

1.3 Stima della deviazione standard short-term

Come possiamo notare dalla formulazione degli indici di capacità, visti nel paragrafo precedente, in tutti c'è una componente della varianza, detta varianza short-term, di seguito riportiamo come effettuare la stima di tale varianza.

La stima della deviazione standard short-term può essere calcolata nei seguenti modi:

- 1) Stima basata sugli scarti tipo:

$$\hat{\sigma}_{ST} = \frac{\bar{s}}{c_4};$$

$$\text{dove } \bar{s} = \frac{\sum_{i=1}^m s_i}{m}; \text{ e } s_i = \sqrt{\frac{\sum_{j=1}^n (x_{ij} - \bar{x}_i)^2}{n-1}}.$$

Per tale stima il numero di osservazioni (n) deve essere maggiore di 2.

2) Stima basata sui range campionari:

$$\hat{\sigma}_{ST} = \frac{\bar{R}}{d_2};$$

dove $R_i = x_{i_{\max}} - x_{i_{\min}}$; e $\bar{R} = \frac{\sum_{j=1}^m R_j}{m}$.

Per tale stima il numero di osservazioni (n) devono essere comprese tra 2 e 10.

3) Stima basata sulle escursioni mobili campionarie

$$\hat{\sigma}_{ST} = \frac{MR(2)}{d_2};$$

dove $MR(2) = |x_j - x_{j-1}|$; e $MR(2) = \frac{\sum_{j=2}^n |x_j - x_{j-1}|}{n-1}$.

Per tale stima si misura la variabilità sulla base di singole misure ossia (n=1).

Le costanti d_2 e c_4 le troviamo tabulate in funzione di n.

1.4 *Lacune degli indici di capacità*

Molteplici sono i contesti applicativi in cui gli indici di capacità sopra riportati non risultano adeguati. La risoluzione dei problemi incontrati è ancora argomento di discussione.

Alcuni dei problemi più noti sono il calcolo degli indici nei casi in cui le osservazioni non hanno una distribuzione normale, e in presenza di componenti di variazione gerarchica o trend.

Alcune discussioni a riguardo le possiamo trovare in Kotz e Lovelace (1998), in Palmer e Tsui (1999) e in Singpurwalla (1998).

Gli aspetti degni di nota tra le impressioni generali e quelle derivate da molte analisi dettagliate possono essere così sintetizzate:

1. L'indice (1.3) è inaffidabile se la proporzione di non conformità attese è visto come l'elemento più importante da trattare, infatti Ruczinski (1996) mostra come gli stessi

valori di tale indice possono essere associati a una vasta gamma di valori della proporzione di non conformità attese.

2. La definizione degli indici di capacità deve essere generalizzata al caso in cui esistono componenti di variazione gerarchica che influiscono sul processo, o in casi in cui al processo partecipi una componente di trend.

In questo ultimo campo, in cui gli studi sono ancora in fase di sviluppo e la letteratura è ancora relativamente limitata, ho realizzato il nucleo centrale della mia tesi.

A seguire il secondo capitolo tratta il metodo ANOVA e “*nested*” ANOVA utilizzato per la stima delle componenti della varianza.

CAPITOLO 2

2.1 Introduzione

In questo secondo capitolo introduco uno strumento molto noto e utile in statistica per la stima delle componenti di variazione nei processi produttivi.

I fattori di variabilità nei processi produttivi possono essere molteplici e riguardano macchine, impianti, attrezzature, materie prime, condizioni di produzione e l'uomo. In questi casi quindi è di fondamentale importanza essere a conoscenza delle concause che contribuiscono alla varianza del processo.

2.2 Analisi della varianza

Lo strumento sicuramente più utilizzato per individuare le componenti significative di variazione e vedere quanto queste incidono sulla variazione totale del processo è il metodo dell'analisi della varianza (ANOVA).

Di seguito riportiamo alcuni modelli per l'analisi della varianza molto utili nei casi che andremo a trattare. In particolare, riportiamo brevemente la descrizione dell'ANOVA annidata, a una e a due vie, ad effetti fissi e ad effetti casuali.

2.2.1 ANOVA con effetti fissi

Il modello di variazione a una via è dato da:

$$y_{ij} = \mu + a_i + e_{ij} \quad (2.1)$$

$i = 1, \dots, m$, $j = 1, \dots, n$, dove m è il numero dei gruppi e n è la dimensione di ciascun gruppo.

Il modello di variazione a due vie è dato da:

$$y_{ijk} = \mu + a_i + b_j + (ab)_{ij} + e_{ijk} \quad (2.2)$$

$i = 1, \dots, n$, $j = 1, \dots, m$, $k = 1, \dots, k$ dove m e k sono i livelli del primo e del secondo fattore e n è il numero delle osservazioni per ciascun livello.

Per i modelli (2.1) e (2.2) le stime delle componenti di variazione sono riportate nella tabella 2.1 e 2.2, rispettivamente:

Fonte	SS	df	Medie dei quadrati
Livelli del fattore	$J \sum a_i^2$	$(I - 1)$	$J \sum a_i^2 / (I - 1)$
Residui	$\sum \sum e_{ij}^2$	$I(J - 1)$	$\sum \sum e_{ij}^2 / (I(J - 1))$
Totale corretto	$\sum \sum y_{ij}^2 - IJ\mu^2$	$IJ - 1$	

Tabella 2.1

Fonte	SS	df	Medie dei quadrati
Righe	$JK \sum a_i^2$	$(I - 1)$	$JK \sum a_i^2 / (I - 1)$
Colonne	$IK \sum b_j^2$	$(J - 1)$	$IK \sum b_j^2 / (J - 1)$
Interazione	$K \sum \sum (ab)_{ij}^2$	$(I - 1)(J - 1)$	$IK \sum b_j^2 / (J - 1)$
Residui	$\sum \sum \sum e_{ijk}^2$	$IJ(K - 1)$	$\sum \sum \sum e_{ijk}^2 / (IJ(K - 1))$
Totale corretto	$\sum \sum y_{ij}^2 - IJ\mu^2$	$IJK - 1$	

Tabella 2.2

2.2.2 ANOVA annidata

Nell'ANOVA annidata tutti i livelli di un fattore vengono combinati con tutti i livelli di un altro fattore, nel caso in cui i livelli di un fattore vengono scelti a caso invece che intenzionalmente si ha un modello ad effetti casuali. Ad esempio, si selezionano dei wafers da un run produttivo e si estraggono delle unità da ciascun wafer. Le unità sono dunque annidate (nested) dentro i wafers e l'effetto di ciascun fattore è casuale.

Si assuma vi siano m wafers, da ciascun wafer si estraggono casualmente n osservazioni. Le misurazioni all'interno di uno stesso wafer tendono ad essere simili. Le osservazioni variano rispetto alla media complessiva a causa dell'effetto del wafer (W) e dell'errore casuale (E) delle osservazioni dentro il wafer. Il modello risultante sarà:

$$y_{ij} = \mu + Wi + \varepsilon_{j(i)} \quad (2.3)$$

$i = 1, \dots, m$, $j = 1, \dots, n$, dove Wi misura l'effetto casuale dell' i -esimo wafer e $\varepsilon_{j(i)}$ misura l'effetto della j -esima misurazione all' interno dell' i -esimo wafer.

La varianza totale delle osservazioni è: $\sigma_T^2 = \sigma_B^2 + \sigma_E^2$.

Dalla relazione $\sigma_{\bar{B}}^2 = \sigma_B^2 + \sigma_E^2 / n$ è possibile ricavare la componente di variazione imputabile al wafer:

$$\hat{\sigma}_B^2 = \hat{\sigma}_{\bar{B}}^2 - \hat{\sigma}_E^2 / n.$$

La media delle n osservazioni per il wafer i è data da: $\bar{x}_i = \frac{\sum_{j=i}^n x_{ij}}{n}$, e sapendo che la

media delle medie è data da: $\bar{\bar{x}} = \frac{\sum_{i=1}^m \sum_{j=i}^n x_{ij}}{nm}$,

la stima della varianza delle medie per wafer diventa:

$$\hat{\sigma}_{\bar{B}}^2 = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{\bar{x}})^2}{m-1}$$

che ci è utile per stimare la componente di variazione imputabile al wafer.

La stima della varianza dell' errore invece è:

$$\hat{\sigma}_E^2 = \frac{\sum_{i=1}^m \sum_{j=1}^n (x_{j(i)} - \bar{x}_i)^2}{m(n-1)}$$

E' inoltre utile verificare se il wafer è una fonte di variazione ignorabile verificando il seguente sistema di verifica d' ipotesi:

$$H_0 : \sigma_B^2 = 0$$

$$H_1 : \sigma_B^2 \neq 0$$

Assumiamo ora che vi siano m lotti (L), da ciascun lotto si estraggono s wafers, (per un totale di $s*m$ wafers), successivamente su ciascun wafer si fanno n misurazioni (per un totale di $s*m*n$ osservazioni).

In questo caso le osservazioni variano rispetto alla media complessiva a causa dell'effetto del lotto, del wafer entro il lotto e dell' errore casuale delle osservazioni entro i wafers.

Il modello risultante sarà:

$$y_{ijk} = \mu + L_i + W_{j(i)} + \varepsilon_{k(ji)} \quad (2.4)$$

$i = 1, \dots, m$, $j = 1, \dots, s$, $k = 1, \dots, n$, dove L_i misura l'effetto casuale dell' i -esimo lotto; $W_{j(i)}$ misura l'effetto casuale del j -esimo wafer annidato dentro l' i -esimo lotto; $\varepsilon_{k(ji)}$ misura l'effetto della k -esima misurazione presa dal j -esimo wafer dentro l' i -esimo lotto.

La varianza totale delle osservazioni è: $\sigma_T^2 = \sigma_L^2 + \sigma_B^2 + \sigma_E^2$,

dove la stima delle componenti di variazione imputabile al lotto e ai wafer entro i lotti sono date da:

$$\hat{\sigma}_L^2 = \hat{\sigma}_{\bar{L}}^2 - \hat{\sigma}_B^2 / s - \hat{\sigma}_E^2 / sn,$$

$$\hat{\sigma}_B^2 = \hat{\sigma}_{\bar{B}}^2 - \hat{\sigma}_E^2 / n,$$

dove:

$$\hat{\sigma}_{\bar{L}}^2 = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{\bar{x}})^2}{m-1},$$

$$\hat{\sigma}_{\bar{B}}^2 = \frac{\sum_{i=1}^m \sum_{j=1}^s (\bar{x}_{ij.} - \bar{x}_{i..})^2}{m(s-1)},$$

e

$$\hat{\sigma}_E^2 = \frac{\sum_{i=1}^m \sum_{j=1}^s \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2}{ms(n-1)},$$

con $\bar{x}_{ij.} = \frac{\sum_{k=1}^n x_{ijk}}{n}$ che è la media per il j -esimo wafer dell' i -esimo lotto e

$\bar{x}_{i..} = \frac{\sum_{j=1}^s \bar{x}_{ij.}}{s} = \frac{\sum_{j=1}^s \sum_{k=1}^n x_{ijk}}{sn}$ che è la media dell' i -esimo lotto.

Le stime delle componenti di variazione associata ai modelli (2.3) e (2.4) sono riportate nelle tabelle 2.3 e 2.4, rispettivamente:

Fonte	SS	df	Medie dei quadrati
Livelli del fattore	$m \sum (X_{i.} - X_{..})^2$	$(m - 1)$	$n \sum (X_{i.} - X_{..})^2 / (m - 1)$
Residui	$\sum \sum (X_{ij} - X_{i.})^2$	$m(n - 1)$	$\sum \sum (X_{ij} - X_{i.})^2 / (m(n - 1))$
Totale corretto	$\sum \sum y_{ij}^2 - mn\mu^2$	$mn - 1$	

Tabella 2.3

Fonte	SS	df	Medie dei quadrati
Righe	$sn \sum (X_{i..} - X_{...})^2$	$(m - 1)$	$sn \sum (X_{i..} - X_{...})^2 / (m - 1)$
Colonne	$mn \sum (X_{ij.} - X_{i..})^2$	$(s - 1)$	$mn \sum (X_{ij.} - X_{i..})^2 / (s - 1)$
Residui	$\sum \sum \sum (X_{ijk} - X_{ij.})^2$	$ms(n - 1)$	$\sum \sum \sum (X_{ijk} - X_{ij.})^2 / (ms(n - 1))$
Totale corretto	$\sum \sum y_{ij}^2 - ms\mu^2$	$msn - 1$	

Tabella 2.4

CAPITOLO 3

3.1 Indici di capacità in presenza di componenti gerarchiche di variazione

L'idea di base nel monitoraggio e nel miglioramento dei processi è di considerare che il processo sia influenzato da un sistema di concause e dobbiamo cercare di capire la natura di tutta la variazione nel processo. Nel 1920 Walter Shewhart iniziò a lavorare su ciò che noi oggi chiamiamo SPC. L'idea di base è quella di ridurre la varianza in uscita dal processo per fare in modo che il processo assomigli a un white noise, ossia un processo soggetto a un costante sistema di cause di variazione.

Nella letteratura un gran numero di autori (Woodall e Thomas 1995, Sparks e Field 2000) ha riconosciuto la fuorviante natura di utilizzare esclusivamente la varianza dei sottogruppi per descrivere l'intera variazione del processo e dunque costruire i limiti di una carta di controllo o per calcolare indici di capacità in presenza di una significativa componente di varianza anche tra i sottogruppi.

Supponendo di avere dei campioni di dimensione n presi ad intervalli discreti equispaziati di Δt cerchiamo di illustrare i differenti effetti della variazione all'interno dei sottogruppi ("high frequency") e della variazione tra i gruppi ("low frequency"). Per effettuare ciò bisogna innanzitutto cogliere l'intera variazione dovuta alle cause comuni che agiscono nel processo e che non possono essere eliminate.

Alla varianza del processo partecipano anche delle cause assegnabili di variazione, tali cause sono eventi imprevisti che potrebbero fare in modo che i prodotti realizzati non siano idonei per l'utilizzo.

Per quanto detto in un gran numero di situazioni pratiche un semplice modello non può essere appropriato. Per processi di produzione con variazione day-to-day o batch-to-batch, tale variazione deve essere considerata parte della variazione naturale del processo. In questi casi è stato dimostrato che la variazione naturale deve includere anche la variazione day-to-day.

Da ciò quindi segue che queste fonti di variazione devono essere incluse nel processo che è in una situazione di controllo statistico.

In tali processi si assume che possono agire quindi due componenti di varianza σ_B^2 e σ_E^2 , anche se possiamo avere anche più di due componenti che determinano la varianza totale dovuta a cause comuni.

I modelli di variazione gerarchica con un effetto casuale dei sottogruppi sono dei buoni modelli per ottenere delle stime robuste delle componenti di variazione indispensabili per una corretta stima degli indici di capacità in presenza di componenti di variazione gerarchica.

Si consideri il modello di variazione (2.3).

La varianza totale delle osservazioni è data da:

$$V[X_{ij}] = V[W_i] + V[\varepsilon_{ij}] = \sigma_B^2 + \sigma_E^2$$

Considerando l'usuale decomposizione ANOVA, la varianza totale può decomporsi nella varianza all'interno dei gruppi e variazione tra i sottogruppi:

$$\sum_{i=1}^N \sum_{j=1}^n (x_{ij} - \bar{x})^2 = \sum_{i=1}^N \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 + \sum_{i=1}^N n(\bar{x}_i - \bar{x})^2$$

e avremo quindi che:

$$s_3^2 = \sum_i \sum_j (x_{ij} - \bar{x}_i)^2 / [N(n-1)]$$

$$s_2^2 = \sum_i (\bar{x}_i - \bar{x})^2 / (N-1)$$

che rappresentano rispettivamente la varianza empirica entro i sottogruppi e la varianza empirica tra i sottogruppi.

Con semplici passaggi otteniamo che:

$$E[s_2^2] = \sigma_B^2 + \sigma_E^2 / n$$

$$E[s_3^2] = \sigma_E^2$$

Con il metodo della stima dei momenti troveremo che:

$$\tilde{\sigma}_E^2 = s_3^2$$

$$\tilde{\sigma}_B^2 = s_2^2 - s_3^2 / n$$

$\tilde{\sigma}_E^2$ e $\tilde{\sigma}_B^2$ sono stimatori non distorti di σ_E^2 e σ_B^2 .

Nelson (1999) ha fornito tre diversi metodi per stimare la deviazione standard del processo necessaria per il calcolo degli indici di capacità.

I metodi citati sono i seguenti:

- 1) Usare le osservazioni individuali con le quali è fatta la carta di controllo,
- 2) Usare le medie dei sottogruppi. Se le medie per sottogruppo sono tutte di dimensione n e l'errore standard di queste medie di sottogruppo è s_a , allora la stima della deviazione standard di quelle individuali è s_a / \sqrt{n} ,
- 3) Usare il range medio dalla carta di controllo.

Quando i sottogruppi razionali racchiudono tutte le cause comuni di variazione, le quali sono riflesse nella variazione all'interno dei sottogruppi, i tre metodi danno essenzialmente lo stesso risultato.

Quando invece la variazione non è tutta riflessa dalla variazione within allora le stime con i tre metodi danno risultati differenti.

I corrispondenti valori attesi sono:

$$E[s_1^2] = \sigma_E^2 + \frac{n(N-1)}{nN-1} \sigma_B^2,$$

$$E[s_2^2] = \sigma_B^2 + \sigma_E^2 / n,$$

$$E[s_3^2] = \sigma_E^2.$$

Analizzando i tre diversi metodi, il primo metodo sembra essere il migliore e s_1^2 sembra il candidato più promettente per una stima grezza della capacità del processo in una situazione bilanciata.

La stima dell'indice sarà quindi:

$$C_p = \frac{USL - LSL}{6s_1},$$

dove:

$$s_1 = \sqrt{s_1^2} = \sqrt{\sigma_B^2 + \sigma_E^2}.$$

Qualora dovessimo invece trovarci in una situazione nella quale l'esperimento non è bilanciato il metodo da utilizzare è il metodo REML (residuals maximum likelihood method).

3.2 Ancora su indici di capacità in presenza di componenti gerarchiche di variazione

In questo paragrafo viene suggerito il calcolo della porzione di variazione (PV) totale dovuta alle cause speciali spiegata dal cambiamento dei parametri controllabili di un processo.

Il PV (Hammett, Majeske, 2007) in sostanza consente di stimare la riduzione percentuale nella variazione dovuta a cause speciali che potrebbe essere raggiunta stabilizzando le variabili del processo, attraverso un consistente set-up e miglioramento nel processo.

Presentiamo di seguito due metodi per il calcolo di tale indice, il modello *process set-up* (PS) e il modello *special cause* (SC). La stima delle componenti della varianza per ambedue i modelli avviene tramite il metodo ANOVA.

Per il modello PS bisogna determinare il numero di batches (b) e il numero di panels o repliche per campione (n), per il modello SC dobbiamo aggiungere il numero di campioni per batches (s).

Nel disegno sperimentale i produttori potrebbero selezionare, il numero di batches, grande abbastanza per catturare la variazione di lungo periodo delle medie di batch. I batches sono da selezionare non consecutivamente e mai di numero inferiore a cinque.

Il numero di campioni per batch deve essere minimo di due osservazioni.

Il numero consigliabile di panels per campione da tre a cinque.

3.2.1 IL MODELLO PROCESS SET-UP

Nel modello PS, X_{ij} rappresenta la j-esima osservazione ($j = 1, \dots, n$) della caratteristica di qualità misurata selezionata per il batch i ($i = 1, \dots, b$).

Il modello è dunque il (2.3) con la componente di batch (b) che sostituisce quella di wafer (w).

Definiamo:

$$\bar{x}_{i.} = \sum_{j=1}^n x_{ij} / n;$$

e

$$\bar{x}_{..} = \sum_{i=1}^b \sum_{j=1}^n x_{ij} / (bn);$$

La somma dei quadrati totale:

$$SST_{PS} = \sum_{i=1}^b \sum_{j=1}^n (x_{ij} - \bar{x}_{..})^2$$

Viene divisa nella somma dei quadrati dovuta all'effetto dei batch:

$$SSB_{PS} = n \sum_{i=1}^b (\bar{x}_{i.} - \bar{x}_{..})^2$$

e nella somma dei quadrati dovuta all'errore:

$$SSE_{PS} = \sum_{i=1}^b \sum_{j=1}^n (x_{ij} - \bar{x}_{i.})^2$$

ovvero:

$$SST_{PS} = SSB_{PS} + SSE_{PS}$$

L'indice R^2 come segue:

$$R_{PS}^2 = SSB_{PS} / SST_{PS}$$

rappresenta la proporzione della variabilità totale spiegata dall'effetto dei batch.

La varianza dovuta ai batch può così essere stimata:

$$\hat{\sigma}_B^2 = \hat{\sigma}_W^2 = [SSB_{PS} / (b-1) - SSE_{PS} / [n(b-1)]] / n$$

Il modello PS classifica la varianza associata ai cambiamenti da batch a batch come una variazione dovuta a cause speciali. Mentre la varianza comune, associata alla variazione part-to-part, è stimata come segue:

$$\hat{\sigma}_E^2 = \hat{\sigma}_e^2 = SSE_{PS} / [n(b-1)] \quad (3.1)$$

La varianza totale risulta:

$$\hat{\sigma}_T^2 = \hat{\sigma}_B^2 + \hat{\sigma}_E^2 \quad (3.2)$$

Si supponga di raccogliere i dati nel set-up del processo e parametri di ingresso (X_1, \dots, X_m), per sviluppare un modello di regressione delle caratteristiche di qualità basato su questi parametri di set-up. Sia y_i rappresenta il valore delle variabili di risposta per $N=bn$ parti dello studio.

Abbiamo quindi un modello della forma:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m + \varepsilon, \quad (3.3)$$

dove X_1, \dots, X_m rappresentano gli m input, le variabili che si ritengono influenzino la caratteristica di qualità y .

Sia \hat{y}_i il valore di y previsto per mezzo del modello di regressione (3.2), la devianza totale di y è:

$$SST_R = \sum_{i=1}^N (y_i - \bar{y})^2$$

La devianza residua è:

$$SSE_R = \sum_{i=1}^N (y_i - \hat{y}_i)^2 .$$

La devianza di y dovuta alla regressione è:

$$SSR_R = SST_R - SSE_R$$

La proporzione della variabilità totale nella caratteristica di qualità spiegata dai parametri di set-up è:

$$R_R^2 = SSR_R / SST_R ,$$

Il tasso PV_{PS} rappresenta la proporzione della variabilità delle medie batch-to-batch associata con cambiamenti nel set-up del processo, e si calcola come segue:

$$PV_{PS} = SSR_R / SSB_{PS} .$$

Stimando poi:

$$\hat{\sigma}_{UE}^2 = SSE_R / [N - (k + 1)]; \quad (3.4)$$

Calcoliamo ora l'indice di capacità (1.1) usando come stima della variazione della caratteristica rispettivamente la (3.1), la (3.2) e la (3.4), ovvero $\hat{\sigma}_E^2$, $\hat{\sigma}_T^2$ e $\hat{\sigma}_{UE}^2$.

In corrispondenza di tali stime della varianza del processo otteniamo gli indici:

$$C_p = \frac{USL - LSL}{6\hat{\sigma}_T} \quad (3.5)$$

dove $\hat{\sigma}_T^2$ è data da (3.2);

Majeske e Hammett (2003) proposero uno stimatore basato sulla variazione SC che può essere stimato usando la variazione part-to-part

$$C_{pp} = \frac{USL - LSL}{6\hat{\sigma}_E} \quad (3.6)$$

dove $\hat{\sigma}_E^2$ è data da (3.1); questo quando si considera che tutta la componente della varianza dovuta alle cause speciali sia stata eliminata;

ed infine

$$C_{PUE} = \frac{USL - LSL}{6\hat{\sigma}_{UE}} \quad (3.7)$$

dove $\hat{\sigma}_{UE}^2$ è data da (3.4).

3.2.2 IL MODELLO SPECIAL CAUSE

Nel modello SC, X_{ijk} rappresenta la k-esima osservazione, $k=1, \dots, n$, del j-esimo campione, $j=1, \dots, s$, estratto dall' i-esimo batch, $i=1, \dots, b$. Usando un modello a due fattori casuali annidando il fattore campione all' interno del fattore batch, il modello risultante è del tipo in (2.4).

La somma dei quadrati totale è:

$$SST_{sc} = \sum_{i=1}^b \sum_{j=1}^s \sum_{k=1}^n (x_{ijk} - \bar{x}_{...})^2$$

La somma dei quadrati dovuta al batch è:

$$SSB_{SC} = ns \sum_{i=1}^b (\bar{x}_{i..} - \bar{x}_{...})^2$$

La somma dei quadrati dovuta alla variazione dei campioni all' interno dei batch è:

$$SSA(B)_{SC} = n \sum_{i=1}^b \sum_{j=1}^s (\bar{x}_{ij.} - \bar{x}_{j..})^2$$

La somma dei quadrati dovuta all'errore è:

$$SSE_{SC} = \sum_{i=1}^b \sum_{j=1}^s \sum_{k=1}^n (x_{ijk} - \bar{x}_{ij.})^2.$$

Quindi la varianza totale può scomporsi nel modo seguente:

$$SST_{sc} = SSB_{SC} + SSA(B)_{SC} + SSE_{SC}$$

dove SSB_{SC} e $SSA(B)$ rappresentano la variabilità attribuibile alle cause speciali.

La proporzione di variabilità dovuta alle cause speciali è data da:

$$R_{SC}^2 = [SSB_{SC} + SSA(B)_{SC}] / SST_{SC}$$

Ora calcoliamo la stima della varianza dovuta ai batch:

$$\hat{\sigma}_L^2 = [SSB_{SC} / (b-1) - SSA(B)_{SC} / (s-1)] / (sn)$$

La stima della varianza dovuta alle differenze in media all'interno dei batch è:

$$\hat{\sigma}_B^2 = \hat{\sigma}_{W(L)}^2 = [SSA(B)_{SC} / (s-1) - SSE / [bs(n-1)]] / sn$$

Per ultima la stima della varianza part-to-part o varianza dovuta a cause comuni:

$$\hat{\sigma}_E^2 = \hat{\sigma}_\epsilon^2 = SSE_{SC} / [ns(b-1)].$$

La varianza totale è quindi stimata da:

$$\hat{\sigma}_T^2 = \hat{\sigma}_L^2 + \hat{\sigma}_B^2 + \hat{\sigma}_E^2.$$

L'indice PV nel caso del modello (2.4) risulta:

$$PV_{SC} = SSR_R / [SSB_{SC} + SSA(B)_{SC}]$$

Tale indice coglie la proporzione della variazione dovuta a cause speciali dovuta dalle proprietà dei materiali e dalle variabili di processo.

CAPITOLO 4

4.1 Simulazione

In questo capitolo si vuole studiare la performance degli indici di capacità (1.2) e (1.4), con $u=0$ e $v=4$ come proposto da Vannman (1995), nel caso in cui esistano componenti significative di variazione gerarchica. In particolare si esplora la distribuzione di tali indici nel caso in cui la variabilità delle medie dei batch sia dovuta solo a cause comuni e nel caso in cui esista una variazione dovuta a cause speciali.

In particolare nel primo caso vengono simulati 1000 volte $b=6$ batches di ampiezza $n=2$ da una normale di media $\mu=20$ e varianza $\sigma_E^2=4$. In questo caso si assume dunque che le medie per batch siano distribuite con media μ e varianza σ_E^2/n . Nel secondo caso le medie dei batches vengono generate da una normale di media $\mu + \delta_i * \sigma_E$ con $\delta = 0.5, 1.5$ e 2.5 . In particolare si considera il caso in cui tutte le medie dei batch subiscano una variazione di $\delta\sigma_E$ e il caso in cui sia solo il 50% delle medie a subire un cambiamento.

Supponiamo che sia $6\sigma_E$ la distanza tra la media μ e i limiti di specifica inferiore e superiore ovvero:

$$LSL = \mu - 6 * \sigma_E, \quad USL = \mu + 6 * \sigma_E$$

e che sia $T = (LSL + USL) / 2$ il target.

Per ogni simulazione, si stimano gli indici di capacità C_{pk} e $C_p(0,4)$, nel caso in cui il processo sia influenzato solo da cause comuni e in quello in cui esistano variazioni significative da batch a batch.

Gli indici C_{pk} e $C_p(0,4)$, nel caso in cui sia nullo lo shift delle medie dei batch, calcolati attraverso la (3.6), rappresentano il massimo valore raggiungibile dagli indici di capacità per tale processo. Per tale caso infatti nel processo partecipano componenti di variazione dovute esclusivamente a cause comuni non eliminabili. Negli altri casi invece in cui le medie dei batch subiscono uno shift piccolo, medio o grande, la formula utilizzata per il calcolo dei due indici è la (3.5). In tali casi nel processo oltre alla variazione dovuta a

cause comuni partecipa anche una componente di variazione dovuta a cause speciali, variazione che può essere eliminata agendo direttamente sul processo di produzione monitorato.

Le Tabelle 4.1.1 e 4.1.2 riportano alcune statistiche di sintesi della distribuzione di frequenza delle stime dell'indice C_{pk} e $C_p(0,4)$ rispettivamente.

Tabella 4.1.1 Statistiche di sintesi della distribuzione delle stime di C_{pk}

Shift	% Medie che variano	Primo quartile	Mediana	Media	Terzo quartile	Deviazioni Standard
0	-	1.0040	1.4830	1.8690	2.2900	1.684
0.5	100%	0.8068	1.1940	1.3840	1.7120	0.806
1.5	100%	0.5691	0.7965	0.9176	1.1350	0.349
2.5	100%	0.4011	0.5704	0.6236	0.77980	0.099
0.5	50%	0.8392	1.1960	1.4140	1.7320	0.739
1.5	50%	0.6670	0.9794	1.1300	1.3900	0.398
2.5	50%	0.4943	0.7093	0.7839	0.9827	0.140

Tabella 4.1.2 Statistiche di sintesi della distribuzione delle stime di $C_p(0,4)$

Shift	% Medie che variano	Primo quartile	Mediana	Media	Terzo quartile	Deviazioni Standard
0	-	0.8991	1.1050	1.1670	1.3990	0.139
0.5	100%	0.8827	1.1030	1.1580	1.3720	0.169
1.5	100%	0.6993	0.8917	0.9973	1.1750	0.514
2.5	100%	0.5564	0.7083	0.8292	0.9123	0.225
0.5	50%	0.8933	1.0990	1.1620	1.3530	0.153
1.5	50%	0.7861	1.0070	1.1050	1.2810	0.222
2.5	50%	0.6625	0.8360	0.9352	1.0930	0.273

I Grafici 4.1.3 e 4.1.4 pongono a confronto la distribuzione delle stime di C_{pk} e di $C_p(0,4)$ al variare degli shift delle medie.

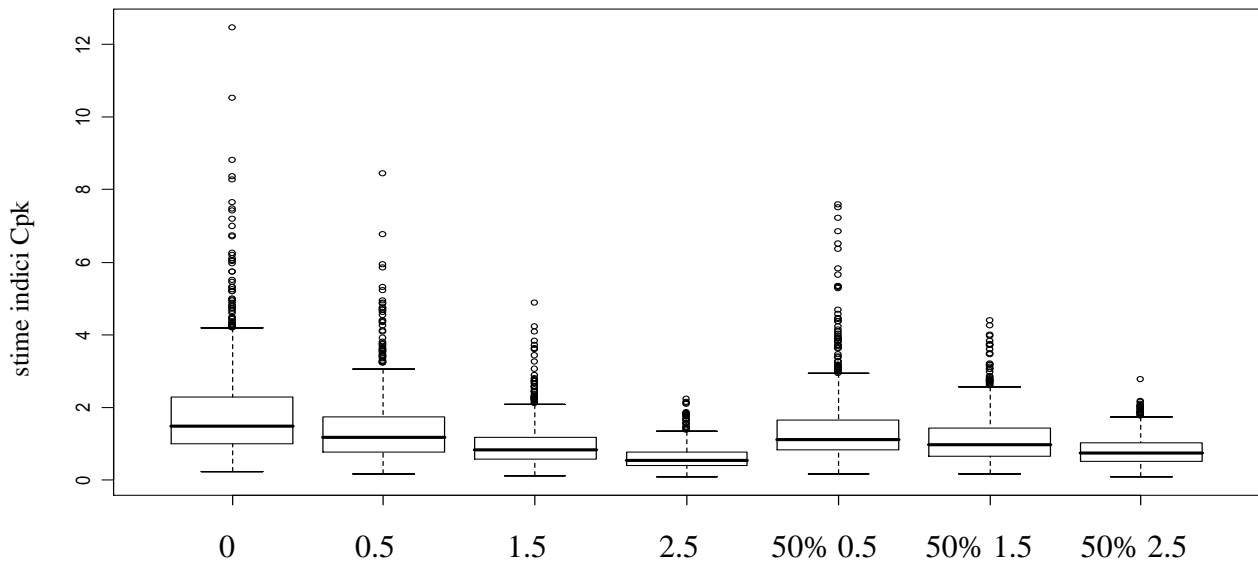


Grafico 4.1.3 Distribuzione delle stime di Cpk in presenza di cause comune e speciali di variazione

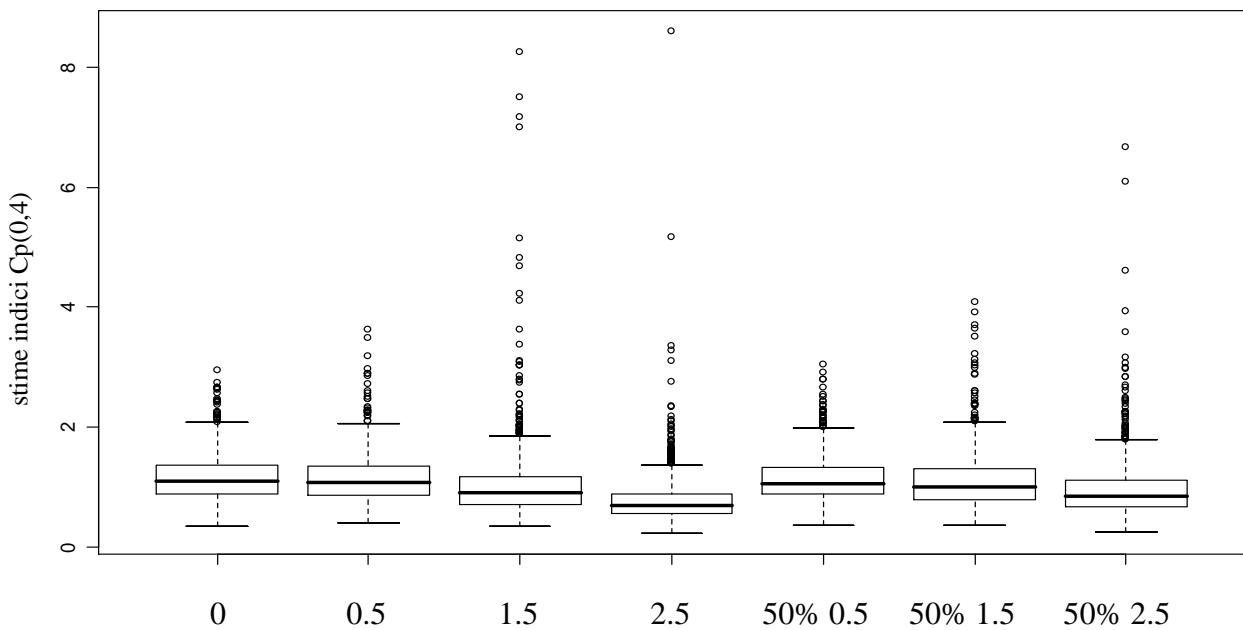


Grafico 4.1.4 Distribuzione delle stime di Cp(0,4) in presenza di cause comune e speciali di variazione

Si osservi che in molti processi produttivi infatti la variazione batch-to-batch rappresenta una caratteristica intrinseca del processo. Pertanto dai risultati ottenuti può essere d'interesse individuare quale margine di variazione nelle medie dei batch produca valori degli indici di capacità ancora accettabili per il cliente. In particolare dalla tabella 4.1.1

possiamo notare che il processo può considerarsi capace nei casi in cui lo shift delle medie dei batch è pari al massimo a 0.5 e nel caso in cui è solo il 50% dei batch a subire uno shift pari a 0.5. Mentre se le medie subiscono uno shift pari a 1.5 siamo in una situazione in cui il processo è marginalmente capace, il valore della mediana è infatti circa 1, mentre in tutti gli altri casi il processo risulta non capace.

Per quanto riguarda invece l'indice $C_p(0,4)$, vedi tabella 4.1.2, il processo può ritenersi capace quando le medie dei batch subiscono uno shift al massimo di 0.5, con shift di 1.5 il processo risulta marginalmente capace, come nel caso precedente, mentre nel caso in cui solo il 50% delle medie dei batch subisce uno shift di 0.5 o 1.5, l'indice fornisce valori leggermente superiori a 1 e dunque consente di classificare ancora il processo come capace.

Si può inoltre notare una sostanziale differenza tra i valori dell'indice C_{pk} e quelli dell'indice $C_p(0,4)$ già per uno shift delle medie per batch pari a 0, infatti le mediane delle distribuzioni risultano rispettivamente 1.4830 e 1.1050, via via che lo shift aumenta la differenza di performance dei due indici si riduce di molto.

Un altro aspetto interessante che possiamo notare è che per valori piccoli degli shift la differenza tra il primo e il terzo quartile per l'indice C_{pk} è molto superiore rispetto alla differenza che possiamo trovare tra gli stessi quartile per la distribuzione dell'indice $C_p(0,4)$. In generale, al variare dello shift, varia di molto il valore dell'indice C_{pk} rispetto all'indice $C_p(0,4)$. In termini di mediana l'indice C_{pk} passa da un valore massimo di 1.4830, per shift pari a 0, a un valore minimo di 0.7093, per shift pari a 2.5. Tale range è molto più contenuto nel caso dell'indice $C_p(0,4)$, infatti in questo caso passiamo da un massimo di 1.1050, per shift pari a zero, a un minimo di 0.8360, per shift pari a 2.5.

In sintesi l'indice C_{pk} sembra molto più sensibile ai salti in media, e quindi all'aumento della varianza batch to batch rispetto all'indice $C_p(0,4)$ che presenta una maggiore variabilità della distribuzione al variare dello shift nelle medie dei batch.

Come atteso, è evidente per entrambi gli indici che all'aumentare del valore degli shift in media il valore degli indici si abbassa in vista dell'aumento della varianza batch to batch e dunque dell'incidenza di questa sulla varianza totale del processo.

In una realtà aziendale in cui gli indici di capacità del processo siano tali da giudicare il

processo non capace si interverrà al fine di ridurre il più possibile la variazione dovuta a cause speciali e avvicinare il valore di tali indici il più vicino possibile alla situazione ottimale rappresentata appunto dalla (3.6). D'altra parte dai risultati di questo studio per simulazione sembrano “tollerabili” shift delle medie tra batch fino a un massimo di 1.5σ .

Conclusioni

Nella trattazione da me svolta ho dato una particolare attenzione esplorando la performance studiata degli indici di capacità in tali contesti allo studio della performance di alcuni indici di capacità nel caso in cui esistano componenti gerarchiche di variazione che influiscono nel processo produttivo.

Dalla simulazione da me realizzata ho messo inoltre a confronto l'indice C_{pk} e l'indice $C_p(0,4)$. In particolare, ho confrontato tali indici nei casi in cui ci sia esclusivamente variazione dovuta a cause comuni e nei casi in cui sono presenti anche componenti di variazione imputabili a cause speciali, monitorando la loro variabilità all'aumentare degli shift delle medie dei batch e quindi all'aumentare della varianza dovuta a cause speciali.

Da tale confronto si evince che, nel caso da me studiato, l'indice $C_p(0,4)$ fornisce valori leggermente migliori rispetto all'indice C_{pk} . L'indice di Vannman sembra infatti meno influenzato dalla presenza di variazioni significative nelle medie dei batch, fornendo risultati meno variabili rispetto all'indice C_{pk} .

Questa trattazione mette in evidenza l'importanza che hanno gli indici di capacità, facendo però anche risaltare le difficoltà che si incontrano nel loro uso; le particolari condizioni che devono essere rispettate per avere una corretta interpretazione e le necessarie modifiche che devono essere apportare nel caso in cui esistano significative fonti di variazione addizionali.

Appendice

Funzione per il calcolo di un campione con shift pari a 0, 0.5, 1.5, 2.5

```
simul1 <-  
function(media=20,sigmaE=2,ngruppi=6,nosservazioni=2,delta=0) {  
  dati<-matrix(data = NA, nrow = ngruppi, ncol = nosservazioni,  
    byrow = TRUE,dimnames = list(c(1:ngruppi),  
c(1:nosservazioni)))  
  
  q<-0  
  z<-0  
  media_campione<-0  
  somma_riga<-c(rep(0,ngruppi))  
  media_righe<-c(rep(0,ngruppi))  
  sigma2bmedio<-0  
  sigmab<-0  
  sigma2b<-0  
  sigma2e<-0  
  
  for (i in 1:ngruppi){  
    for(j in 1:nosservazioni){  
      Eij<-rnorm(1,0,sigmaE)  
      dati[i,j]<-media+delta*Eij+(Eij^2/nosservazioni)  
      somma_riga[i]<-somma_riga[i]+dati[i,j]  
    }  
    media_righe[i]<-somma_riga[i]/nosservazioni  
  }  
  if (delta==0){  
#stima media campione  
    media_campione<-mean(media_righe)  
  
    for(i in 1:ngruppi){  
      for(j in 1:nosservazioni){  
        q<-q+((dati[i,j]-media_righe[i])^2)/(nosservazioni-1)  
      }  
    }  
#stima di sigma2e  
    sigma2e<-q/ngruppi  
  }  
  else{  
#stima media campione  
    media_campione<-mean(media_righe)  
  
    for(i in 1:ngruppi){  
      for(j in 1:nosservazioni){  
        q<-q+((dati[i,j]-media_righe[i])^2)/(nosservazioni-1)  
      }  
    }  
  }  
}
```

```

z<-z+(media_righe[i]-media_campione)^2
}

#stima di sigma2e
sigma2e<-q/ngruppi

#stima di sigma2bmedio
sigma2bmedio<-z/(ngruppi-1)

#stima di sigma2b
sigma2b<-sigma2bmedio-(sigma2e/nosservazioni)}

list(campione=dati,mu=media_campione,sigma2e=sigma2e,sigma2b=sigma
2b,sigma2bmedio=sigma2bmedio)
}

```

Funzione per il calcolo di 1000 campioni con shift pari a 0, 0.5, 1.5, 2.5

```

simula1000<-
function(media=20,sigmaE=2,ngruppi=6,nosservazioni=2,delta=0,ncamp
=1000){
  osservazioni<-list()
  medie<-c(rep(0,ncamp))
  s2b <-c(rep(0,ncamp))
  s2e <-c(rep(0,ncamp))
  sb <-c(rep(0,ncamp))
  se <-c(rep(0,ncamp))
  s2bmedio<-c(rep(0,ncamp))
  for(i in 1:ncamp){
    campione<-
simula1(media,sigmaE,ngruppi,nosservazioni,delta)
    medie[i]<-campione$mu
    s2b[i]<-campione$sigma2b
    s2e[i]<-campione$sigma2e
    if (campione$sigma2b>0)
    sb[i]<-sqrt(campione$sigma2b)
    else
    sb[i]<-0
    se[i]<-sqrt(campione$sigma2e)
    s2bmedio[i]<-campione$sigma2bmedio
    osservazioni[[i]]<-campione}

list(medie=medie,sigma2b=s2b,sigmab=sb,sigma2e=s2e,sigmae=se,sigma
2bmedio=s2bmedio)}

```

Funzione per il calcolo di un campione con il 50% delle osservazioni con shift di 0.5, 1.5, 2.5

```
simulalmisto <-  
function(media=20,sigmaE=2,ngruppi=6,nosservazioni=2,delta=0.5) {  
  dati<-matrix(data = NA, nrow = ngruppi, ncol = nosservazioni,  
    byrow = TRUE,dimnames = list(c(1:ngruppi),  
c(1:nosservazioni)))  
  
  q<-0  
  z<-0  
  media_campione<-0  
  somma_riga<-c(rep(0,ngruppi))  
  media_righe<-c(rep(0,ngruppi))  
  sigma2bmedio<-0  
  sigmab<-0  
  sigma2b<-0  
  sigma2e<-0  
  
  E11<-rnorm(1,0,sigmaE)  
  dati[1,1]<-media+(E11^2/nosservazioni)  
  E12<-rnorm(1,0,sigmaE)  
  dati[1,2]<-media+(E12^2/nosservazioni)  
  
  somma_riga[1]<-dati[1,1]+dati[1,2]  
  media_righe[1]<-somma_riga[1]/nosservazioni  
  
  E21<-rnorm(1,0,sigmaE)  
  dati[2,1]<-media+delta*E21+(E21^2/nosservazioni)  
  
  E22<-rnorm(1,0,sigmaE)  
  dati[2,2]<-media+delta*E22+(E22^2/nosservazioni)  
  
  somma_riga[2]<-dati[2,1]+dati[2,2]  
  media_righe[2]<-somma_riga[2]/nosservazioni  
  
  E31<-rnorm(1,0,sigmaE)  
  dati[3,1]<-media+delta*E31+(E31^2/nosservazioni)  
  E32<-rnorm(1,0,sigmaE)  
  dati[3,2]<-media+delta*E32+(E32^2/nosservazioni)  
  
  somma_riga[3]<-dati[3,1]+dati[3,2]  
  media_righe[3]<-somma_riga[3]/nosservazioni  
  
  E41<-rnorm(1,0,sigmaE)  
  dati[4,1]<-media+(E41^2/nosservazioni)  
  E42<-rnorm(1,0,sigmaE)  
  dati[4,2]<-media+(E42^2/nosservazioni)
```

```

somma_riga[4]<-dati[4,1]+dati[4,2]
media_righe[4]<-somma_riga[4]/nosservazioni

E51<-rnorm(1,0,sigmaE)
dati[5,1]<-media+(E51^2/nosservazioni)
E52<-rnorm(1,0,sigmaE)
dati[5,2]<-media+(E52^2/nosservazioni)

somma_riga[5]<-dati[5,1]+dati[5,2]
media_righe[5]<-somma_riga[5]/nosservazioni

E61<-rnorm(1,0,sigmaE)
dati[6,1]<-media+delta*E61+(E61^2/nosservazioni)
E62<-rnorm(1,0,sigmaE)
dati[6,2]<-media+delta*E62+(E62^2/nosservazioni)

somma_riga[6]<-dati[6,1]+dati[6,2]
media_righe[6]<-somma_riga[6]/nosservazioni

#stima media campione
media_campione<-mean(media_righe)

for(i in 1:ngruppi){
  for(j in 1:nosservazioni){
    q<-q+((dati[i,j]-media_righe[i])^2)/(nosservazioni-1)
  }
  z<-z+(media_righe[i]-media_campione)^2
}

#stima di sigma2e
sigma2e<-q/ngruppi

#stima di sigma2bmedio
sigma2bmedio<-z/(ngruppi-1)

#stima di sigma2b
sigma2b<-sigma2bmedio-(sigma2e/nosservazioni)
list(campione=dati,mu=media_campione,sigma2e=sigma2e,sigma2b=sigma
2b,sigma2bmedio=sigma2bmedio
}

```

Funzione per il calcolo di 1000 campioni con il 50% delle osservazioni con shift di 0.5, 1.5, 2.5

```
simula1000misto<-  
function(media=20,sigmaE=2,ngruppi=6,nosservazioni=2,delta=0.5,ncamp=1000){  
  osservazioni<-list()  
  
  medie<-c(rep(0,ncamp))  
  s2b  <-c(rep(0,ncamp))  
  s2e  <-c(rep(0,ncamp))  
  sb   <-c(rep(0,ncamp))  
  se   <-c(rep(0,ncamp))  
  s2bmedio<-c(rep(0,ncamp))  
  for(i in 1:ncamp){  
    campione<-  
simulalmisto(media,sigmaE,ngruppi,nosservazioni,delta)  
    medie[i]<-campione$mu  
    s2b[i]<-campione$sigma2b  
    s2e[i]<-campione$sigma2e  
    if (campione$sigma2b>0)  
    sb[i]<-sqrt(campione$sigma2b)  
    else  
    sb[i]<-0  
    se[i]<-sqrt(campione$sigma2e)  
    s2bmedio[i]<-campione$sigma2bmedio  
    osservazioni[[i]]<-campione}  
  
list(medie=medie,sigma2b=s2b,sigmab=sb,sigma2e=s2e,sigmae=se,sigma2bmedio=s2bmedio)}
```

Calcolo dei limiti di specifica

```
LSL<-20-6*2  
USL<-20+6*2  
T<-(LSL+USL)/2
```

Funzione per il calcolo degli indici Cpk e Cp(0,4)

```
CalcolaIndici<-function(medie, sb, se, s2b, s2e, LSL, USL, T, ncamp=1000) {  
  
  Cpk<-c(rep(0, ncamp))  
  d<-c(rep(0, ncamp))  
  M<-c(rep(0, ncamp))  
  Cp04<-c(rep(0, ncamp))  
  val1<-c(rep(0, ncamp))  
  val2<-c(rep(0, ncamp))  
  den<-c(rep(0, ncamp))  
  sigmat<-c(rep(0, ncamp))  
  rad<-c(rep(0, ncamp))  
  for(i in 1:ncamp)  
  {  
    val1[i]<-(medie[i]-LSL)/(3*(sb[i]+se[i]))  
    val2[i]<-(USL-medic[i])/(3*(sb[i]+se[i]))  
    Cpk[i]<-min(val1[i], val2[i])  
    d[i]<-(USL-LSL)/2  
    sigmat[i]<-s2b[i]+s2e[i]  
    rad[i]<-sigmat[i]+4*(medie[i]-T)  
    if(rad[i]<0){  
      Cp04[i]<-0  
    }  
    else{  
      den[i]<-3*sqrt(sigmat[i]+4*(medie[i]-T))  
      Cp04[i]<-d[i]/den[i]  
    }  
  }  
  list(Cpk=Cpk, Cp04=Cp04)  
}
```

Bibliografia

1. Burdick R. K., Borror C. M. e Montgomery D.C. (2003). "A review of methods for measurement systems capability analysis", *Journal of Quality Technology*, vol. 35 n°4, pp 342-353
2. Gijo E. V. (2005). "Improving process capability of manufacturing process by application of statistical techniques", *Quality Engineering*, vol. 17 n°2, pp 309-315.
3. Kotz S. e Johnson N. L. (2002). "Process capability indices- a review, 1992-2000 ", *Journal of Quality Technology*, vol. 34 n°1, pp 2-19.
4. Majeske K. D. e Hammett P.C. (2003). "Identifying sources of variation in sheet metal stamping", *The International Journal of Flexible Manufacturing System*, vol.15 n°1, pp 5-18.
5. Majeske K. D. e Hammett P.C. (2007). "Quantifying the variance contribution of special cause mean shifts in reconfigurable batch manufacturing processes", *International Journal of Production Research*, vol. 45 n°12, pp 2799-2817.
6. Thyregod P., Melgaard H. e Iwersen J. (2003). "On monitoring processes and assessing process capability under a hierarchical model, Part 1", *Frontiers in Statistical Quality Control*, pp 150-167.
7. Thyregod P., Melgaard H. e Iwersen J. (2003). "On monitoring processes and assessing process capability under a hierarchical model, Part 2", *Frontiers in Statistical Quality Control*, pp 168-193.
8. Xie M., Tsui K. L., Goh T. N., Cai D. Q. (2002). "Process capability indices for a regularly adjusted process", *International Journal of Production Research*, vol. 40 n°10, pp 2367-2377.

Ringraziamenti

Al termine di questa tesi desidero fare alcuni ringraziamenti in particolare alla mia famiglia che mi ha permesso di intraprendere la carriera universitaria e mi hanno sostenuto nei momenti di difficoltà.

Desidero ringraziare la professoressa Capizzi che mi ha aiutato e supportato nella stesura della tesi.

Inoltre un grazie speciale a tutti i miei amici che mi sono sempre stati vicino in particolare Mirko per l' aiuto reciproco che ci siamo dati in questi tre anni.

Questa tesi è dedicata a una persona in particolare, mio nonno.

