

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE



TESI DI LAUREA

ANALISI E CONFRONTI DI INDICI DI CAPACITA'
IN CASO DI NON – NORMALITA'

Relatore: Ch.mo prof. SILVANO BORDIGNON

Laureando: ALESSANDRO CAMPORESE

ANNO ACCADEMICO 2002 - 2003

Alle tre persone più importanti della
mia vita:
Giuseppe, Donatella e Mariastefania

INDICE

INTRODUZIONE	pag. 1 – 2
CAP.1 Introduzione agli Indici di Capacità.	
1.1. Cos'è lo studio della Capacità di un processo	pag. 3 – 5
1.2. Introduzione agli Indici di Capacità	pag. 6 – 8
1.3. Struttura di base degli Indici di Capacità e primi problemi riguardanti il loro uso	pag. 9 – 11
1.4. Il dibattito sull'uso o no degli Indici di Capacità	pag. 12 – 13
CAP.2 Gli Indici di Capacità C_p e C_{pk}.	
2.1. Introduzione	
2.1.1 Sono il C_p e il C_{pk} un ostacolo per il continuo miglioramento? Esiste una loro alternativa?	pag. 15 – 18
2.2. L'indice di Capacità C_p	
2.2.1. Definizione di C_p	pag. 19 – 22
2.2.2. Caratteristiche di C_p	pag. 22 – 25
2.3. L'Indice di Capacità C_{pk}	
2.3.1 Definizione di C_{pk}	pag. 26 – 30
2.3.2 Caratteristiche di C_{pk}	pag. 30 – 32
CAP.3 L'Indice C_{pm} e altri suoi indici associati.	
3.1. Introduzione	pag. 33 – 35
3.2. L'indice di Capacità C_{pm}	
3.2.1. Definizione di C_{pm}	pag. 36 – 38
3.2.2. Caratteristiche di C_{pm}	pag. 38 – 39
3.3. L'Indice di Capacità C_{pk}	
3.3.1. I tre Indici di Capacità nel caso di tolleranze asimmetriche: C_{pm}^* , C_{pm}^+ e C_{pp}''	pag. 40 – 41
3.3.2 L'Indice $C_{pm}(a)$	pag. 42
3.3.3. L'Indice C_{pmk}	pag. 42 – 43

CAP.4 Indici di Capacità per distribuzioni Non-Normali.

4.1.	Introduzione	pag. 45 – 47
4.2.	Gli effetti della Non-Normalità	
4.2.1.	Introduzione	pag. 48 – 50
4.2.2.	Il caso della Normale contaminata	pag. 50 – 52
4.2.3.	Il caso della distribuzione di Edgeworth	pag. 52 – 53
4.2.4.	Altri casi	pag. 53 – 54
4.3.	Trasformazione dei dati distribuiti non-normalmente in dati distribuiti normalmente: "il metodo di Johnson".	
4.3.1.	Introduzione	pag. 55 – 56
4.3.2.	Le trasformazioni di Johnson	pag. 56 – 57
4.3.3.	Procedura di selezione e di stima dei parametri	pag. 57 – 58
4.3.3.1.	Procedura di selezione	pag. 58 – 59
4.3.3.2.	Stima dei parametri	pag. 59 – 61
4.3.4.	Test per verificare la normalità dei dati trasformati	
4.3.4.1.	Test di Shapiro-Wilk	pag. 61 – 62
4.3.4.2.	Test di Kolmogorov-Smirnov	pag. 62 – 63
4.3.5.	Classificazione delle più importanti distribuzioni secondo l'approccio di Johnson e relative conclusioni	pag. 63 – 64
4.4.	Metodo di Clements, Metodo di Johnson-Kotz-Pearn e delle distribuzioni libere	
4.4.1.	Introduzione	pag. 65 – 66
4.4.2.	Il metodo di Clements	pag. 67 – 68
4.4.2.1.	Uso della curva di distribuzione di Pearson	pag. 68 – 69
4.4.3.	Il metodo di Johnson-Kotz-Pearn	pag. 70
4.4.3.1.	Differenza tra il metodo di Clements e quello di Johnson-Kotz-Pearn	pag. 71
4.4.3.2.	L'approccio di Munechika	pag. 71 – 72
4.4.4.	Indici di Capacità per le distribuzioni 'libere'	pag. 72
4.5.	Altri fondamentali Indici di Capacità per le distribuzioni Non-Normali e Asimmetriche	
4.5.1.	Un indice di Capacità flessibile $C_{j_{kp}}$	pag. 73 – 74
4.5.2.	Indice C_s di Wright	pag. 74 – 75
4.5.2.1	Lo stimatore di C_s	pag. 75 – 77

4.5.3.	Indici di Capacità basati sulla deviazione standard pesata (WSD)	
4.5.3.1	Il metodo della deviazione standard pesata	pag. 77 – 78
4.5.3.2	L'Indice C_p basato sul metodo WSD	pag. 79 – 80
4.5.3.3	L'Indice C_{pk} basato sul metodo WSD	pag. 80 – 81
4.5.3.4	L'Indice C_{pm} e C_{pm}^* basato sul metodo WSD	pag. 81 – 82
4.5.3.5	Performance degli Indici di Capacità WSD	pag. 82 – 83
4.5.4.	Indici di Capacità basati sul metodo della varianza pesata (WV)	pag. 83 – 84
4.5.5.	L'Indice generalizzato $C_{Np}(u, v)$	
4.5.5.1.	Definizione di $C_{Np}(u, v)$	pag. 84 – 86
4.5.5.2.	Stima di $C_{Np}(u, v)$	pag. 86

CAP.5 Le principali distribuzioni Non - Normali.

5.1.	La distribuzione Gamma ($\Gamma(p, \lambda)$)	pag. 87 – 89
5.2.	La distribuzione Chi – Quadro (χ^2_u)	pag. 90 – 93
5.3.	La distribuzione t di Student (t_n)	pag. 94 – 96
5.4.	La distribuzione F di Fischer – Snedecor ($F_{m, n}$)	pag. 97 – 99
5.5.	La distribuzione Lognormale ($\ln(x)$)	pag. 100 – 102
5.6.	La distribuzione Esponenziale ($\exp(\lambda)$)	pag. 103 – 105

CAP.6 Confronto tra gli Indici di Capacità tramite simulazione

6.1.	Metodo di ricerca	pag. 107
6.2.	Analisi della distribuzione Gamma(p, λ)	pag. 108 – 113
6.3.	Analisi della distribuzione Chi-Quadro (χ^2_u)	pag. 114 – 119
6.4.	Analisi della distribuzione t di Student (t_n)	pag. 120 – 125
6.5.	Risultati finali e analisi finale	pag. 126 – 128

CONCLUSIONI pag. 129

APPENDICE pag. 131 – 136

BBLIOGRAFIA pag. 137 – 140

INTRODUZIONE

Con l'avvento degli anni '80 comincia a svilupparsi all'interno delle grandi imprese il processo del Controllo Statistico della Qualità. Questo processo, sviluppatosi in modo esponenziale con il passare degli anni fino a raggiungere il suo picco in questi ultimi anni, è diventato parte integrante di tutto il mondo imprenditoriale, dalla medie-piccole imprese alle grandi imprese. Infatti, per ogni impresa, diventa sempre più importante soddisfare quelle che sono le esigenze e le richieste del cliente affinché questo possa essere soddisfatto del proprio operato. Lo studio della Capacità di un processo risolve quindi i problemi, trovando ed eliminando le cause cosiddette "speciali" e decidendo quando l'azione del management è necessaria per ridurre la variabilità del processo sino a livello desiderato. Per fare questo ogni azienda deve quindi monitorare i propri processi produttivi e controllarli, controllo che viene svolto attraverso degli Indici di Capacità.

Capacità per l'appunto letteralmente significa "l'abilità nel compiere un incarico o ottenere un risultato che viene definito dalle specifiche richieste del cliente", e i due fondamentali e più conosciuti Indici di Capacità sono il C_p e il C_{pk} . Questi due Indici però richiedono delle condizioni molto specifiche, e precisamente che i dati del processo siano distribuiti normalmente, siano incorrelati e che il processo sia sotto controllo.

Molto spesso però, all'interno dei processi produttivi aziendali, queste tre condizioni difficilmente convivono insieme e in particolare è difficile trovare dei processi che si distribuiscono in modo perfettamente Normale. È a causa di questa difficoltà che ho pensato di sviluppare il mio elaborato attorno agli Indici di Capacità per le distribuzioni Non – Normali. La mia tesi si sviluppa in due parti, una parte teorica e una parte pratica. La parte teorica è formata dai primi 5 Capitoli mentre il Capitolo 6 è la parte pratica formata da una analisi attraverso una simulazione.

Per quanto riguarda la parte teorica, ho iniziato con una prima parte introduttiva formata dai Capitoli 1, 2, 3.

Nel Capitolo 1 ha fatto una piccola introduzione riguardante quello che è il concetto di Capacità e degli Indici di Capacità di base, trattando anche quello che è il dibattito sorto tra i sostenitori della teoria e quelli della pratica relativamente all'uso o no di questi Indici.

Nel Capitolo 2 ho analizzato il C_p ed il C_{pk} con le loro relative caratteristiche e nel Capitolo 3 ho analizzato il C_{pm} con i suoi rispettivi Indici associati.

È nel Capitolo 4 che si trova il fulcro di tutta la parte teorica, è in questo che sviluppo i principali Indici di Capacità per le distribuzioni Non – Normali.

All'interno di questo Capitolo vengono trattati diversi metodi che sono stati analizzati da vari studiosi, e precisamente:

- Kotz e Johnson (1993) hanno sviluppato degli Indici robusti attraverso il metodo di Clement's, il metodo di Johnson-Pearn-Kotz e i gli Indici di Capacità con distribuzione libera.
- Wright (1995) ha sviluppato l'Indice C_s , derivante da una modifica fatta al denominatore del C_{pmk} in cui si aggiunge un fattore di correzione per l'asimmetria.
- Choi e Bai (1996) ha proposto un metodo basato sulla deviazione standard pesata e sulla varianza standard pesata, cercando così di ottenere un Indice che tiene conto dei gradi di asimmetria tenendo in considerazione separatamente la deviazione standard sopra e sotto la media.
- Vamman ha proposto un Indice generalizzato per dare così una generalizzazione ai quattro Indici C_p , C_{pk} , C_{pm} e C_{pmk} .

Dopo questa analisi teorica dettagliata ho pensato di ricercare, attraverso una simulazione delle principali distribuzioni Non – Normali (distribuzioni descritte all'interno del Capitolo 5), il metodo che risulta essere più corretto e adatto tra tutti quelli analizzati.

Per fare questo, come detto, nel Capitolo 6 ho fatto una simulazione attraverso l'uso di R di tre distribuzioni Non – Normali sotto delle determinate condizioni, e precisamente la Gamma, la Chi – Quadro e la t di Student. Ho fatto questa scelta perché ho pensato di confrontare due distribuzioni Non – Normali e asimmetriche e una Non – Normale ma simmetrica. Ho usato un particolare metodo di simulazione dal momento che ho mantenuto costante la media e la varianza mentre ho fatto variare il coefficiente di asimmetria, di curtosi e la numerosità campionaria, ottenendo così dei risultati interessanti che mi hanno portata a delle conclusioni su quali sono i metodi più corretti.

Per quanto riguarda le distribuzioni Non – Normali e asimmetriche il metodo migliore è quello delle deviazioni standard pesate anche se con una differenza percentuale media attorno al 12%, mentre per le distribuzioni Non – Normali e simmetriche il metodo migliore è quello di Clements anche se questo ha una differenza media percentuale attorno al 13%.

CAPITOLO 1

INTRODUZIONE AGLI INDICI DI CAPACITÀ.

1.1 Cos'è lo studio della Capacità di un processo.

Cominciamo con il porci una domanda: “Perché dovremmo interessarci allo studio della Capacità di un processo? “. La chiave per rispondere a questa domanda può essere ricavata da un articolo di Ben Gunter (1997) scritto per il Quality Progress, nel quale si sottolinea come il bisogno dell'azienda sia quello di capire quanto importante e necessario sia il continuo miglioramento per sé stessa.

La frase “continuo miglioramento della qualità” viene interpretato come il continuo miglioramento nella qualità dei processi produttivi dell'azienda. Infatti, nella mente di molta gente, il concetto di “miglioramento nella qualità “ viene associato alla qualità del prodotto e si ritiene che questo miglioramento comporti per esempio maggiori caratteristiche per i prodotti, l'uso di materiali di lusso, maggiori servizi, uno sviluppo più veloce venendo quindi a costare molto il prodotto e richiedendo molto tempo per la sua produzione. Sfortunatamente però questo modo di vedere e pensare la qualità crea una visione sbagliata.

Infatti quando si tratta e si parla di “continuo miglioramento della qualità” ci si riferisce al miglioramento dei processi produttivi che producono i prodotti ed erogano i servizi, miglioramento che non costa ma fa risparmiare e quindi risparmio che proviene dal ridurre quello che per una azienda viene considerato “superfluo” e che alla fine viene a costare.

Allo stesso tempo anche Shewhart cercò di studiare questo concetto di “miglioramento” e la sua fondamentale osservazione fu quella che tutti i processi producono dei risultati che variano e che questa variabilità può essere divisa in due classi:

- a) Cause accidentali
- b) Causa identificabili

Con questo quindi Shewhart cerca di raggruppare da un lato, una serie di cause cosiddette "accidentali" o "non identificabili" e di singola scarsa influenza che essendo inevitabili, se non a costo di inevitabili revisioni, propongono un comportamento del processo produttivo da considerare come tipico del processo stesso e quindi a carattere casuale, mentre dall'altro lato si tende ad eliminare quell'insieme di cause che producono effetti di più rilevante peso, potendo variare l'equilibrio dei processi produttivi essendo esse, sia tecnicamente che economicamente, eliminabili.

Nei processi produttivi delle varie aziende le cause cosiddette "identificabili" ed "eliminabili" sono per esempio la taratura delle macchine, il consumo delle attrezzature o il controllo imperfetto della temperatura, gli errori di immissione dati, gli errori di spedizione ed inadeguate istruzioni nelle procedure e nei programmi dei computer. Tutte queste cause "identificabili" e determinabili possono essere eliminate richiedendo delle piccole spese che successivamente però producono dei benefici.

Le cause "accidentali" invece possono essere per esempio le rotture ed usure che capitano di solito, variazioni dovute all'umidità, misurazioni imprecise, errori di base nell'immissione dati o modifiche del tempo di risposta di un computer dovute al troppo caricamento di informazioni e dati da parte di chi lo usa. Queste cause di variabilità, essendo parte integrante del sistema produttivo stesso, possono essere ridotte e corrette solamente dalla maggior manutenzione, dal comprare migliori strumenti e macchine di lavoro, dal potenziamento degli hardware e software dei computer. Anche questi miglioramenti costano ma sicuramente alla fine, proprio come l'eliminazione delle cause "identificabili", danno dei benefici a lungo tempo.

Dal momento che la variabilità e le sue cause non potranno mai totalmente essere eliminate, il controllo e la diminuzione della variabilità sono la chiave per ottenere una buona Qualità del processo produttivo. Infatti se un processo è controllabile allora diventa predicabile e i prodotti soddisferanno le richieste del cliente.

È per questo che viene definita "l'Analisi della Capacità di un processo", intendendo con la parola Capacità di un processo semplicemente la variabilità che un processo produttivo rivela quando è influenzato solo dalle cause "accidentali".

Secondo l'AT&T Statistical Quality Control Handbook (1995), la Capacità di un processo è "la serie predicabile degli effetti prodotti da un processo quando si ammette di operare senza l'interferenza di cause esterne...la miglior distribuzione che può essere mantenuta in controllo statistico per un periodo considerato sotto una serie di date condizioni".

Come dice questo libro dunque, lo studio della capacità di un processo risolve i problemi trovando ed eliminando le cause cosiddette "speciali" e decidendo quando l'azione del management è necessaria per ridurre la variabilità del processo sino al livello desiderato.

Gli ingegneri e manager della Qualità sono infatti spesso occupati con delle decisioni riguardanti la risoluzione di alcuni problemi, come quelli se accettare o scartare l'insieme di output del processo produttivo, se intervenire o no nel processo produttivo stesso e se rivedere o no le decisioni da loro prese riguardo il possibile intervento.

"Questo tipo di decisioni molto spesso - secondo Byron e Irony - si basano sulla stima percentuale dei prodotti non conformi in un insieme o lotto di output", e sempre secondo loro "Molte tecniche che aiutano a prendere decisioni, ottenendo l'effetto desiderato sul processo grazie l'analisi della capacità del processo, sono ora utilizzabili ed accentuano le decisioni prese basandosi solamente sui valori previsti".

Quindi, come si è sottolineato, il risultato di una analisi accurata riguardante la Capacità di un processo concede accurate informazioni su quelle decisioni che il management di una azienda dovrà prendere per fare in modo che le specifiche richieste dal cliente siano rispettate portando così ad una sua soddisfazione

1.2. Introduzione agli Indici di Capacità (Process Capability Indices).

In prima istanza definiamo cosa letteralmente significa la parola Capacità: "L'abilità nel compiere un incarico o ottenere un risultato che viene definito dalle specifiche richieste dal cliente".

Tutto quindi comincia dalla definizione e da una domanda di base alla quale per fortuna si è sempre trovata risposta, che veniva e che viene tuttora fatta all'interno di ogni organizzazione orientata al Controllo della Qualità: "Siamo capaci di soddisfare le esigenze del nostro cliente?". A questa domanda, che sembra piuttosto semplice in superficie, le difficoltà incontrate hanno portato ad una valanga di ricerche e dibattiti in tutto il mondo.

Il primo leader nell'uso del Controllo della Qualità e nel continuo miglioramento verso la qualità fu Bell Laboratories, che contemporaneamente fu coadiuvato da Walter A. Shewhart (1931), considerato il pioniere del Controllo della Qualità e il quale introdusse le carte di controllo Shewhart, carte che sono tuttora in uso e che continuano a servire come base del Controllo Statistico della Qualità. Queste carte furono uno dei primi strumenti sviluppati dalla teoria statistica per monitorare la variabilità dei processi produttivi. Variabilità che viene considerata l'aspetto condizionante per l'ottenimento di un output perfetto da ogni processo produttivo e nella quale si focalizza l'attenzione per capire quali siano le sue caratteristiche, apprezzando quelle che sarebbero controllabili.

Il genio e la bravura di Shewhart furono sottolineate dall'aver trovato quali procedure potevano essere utilizzate per eliminare le cause identificabili sino a riuscire a pervenire ai limiti di capacità della varianza di un processo (o ancora meglio per essere precisi riuscì quasi a pervenire ai limiti di capacità della varianza dal momento che il completo successo raramente è possibile) .

Per questo motivo, nel suo libro intitolato "The American who taught the Japanies about Quality", il dottor Deming (1955) spiega ai manager delle aziende quale e come sia il nuovo business e quali sia il punto cruciale del nuovo management, documentando e trovando come il principale male del management americano sia dovuto alla confusione che proviene dal non saper riconoscere le cause di variabilità accidentali da quelle identificabili, ottenendo così un risultato inefficiente ed inefficace per le proprie aziende.

Il dottor Deming quindi, il cui lavoro venne ignorato dagli americani, fu invitato dall'Unione degli Scienziati ed Ingegneri Giapponesi ad insegnare alle aziende in via di sviluppo il significato degli Indici di Capacità e il significato del concetto di Qualità Totale e, grazie alla sua guida, il sistema di Controllo della Qualità Giapponese divenne uno dei più raffinati al mondo di quegli anni, scoprendo che la loro unica possibilità di crescita doveva essere legata alla produzione di beni di alta Qualità, cosa alla quale gli americani difficilmente pensavano.

In un primo momento dunque i Giapponesi scoprirono che qualcosa era stato trascurato, infatti erano riusciti a portare i loro processi produttivi sotto controllo ma non riuscivano a misurare la Capacità di questi processi nel produrre beni accettabili. A sottolineare questa situazione viene dimostrato come, per mettere a confronto la distribuzione dell'output di una azienda con le specifiche del processo, le aziende stesse usassero semplicemente gli istogrammi, le carte di controllo, usando questi metodi per assicurarsi che il processo fosse sotto controllo.

La soluzione a questo problema si trovò negli indici di capacità Giapponesi, che sono:

- a) C_p = Capacità del processo
- b) K = Posizione del processo in relazione alla specificazione della media
- c) C_{pk} = Posizione del processo 6σ in relazione alla media
- d) C_{pl} = Posizione del processo 6σ in relazione al limite inferiore
- e) C_{pu} = Posizione del processo 6σ in relazione al limite superiore

Il concetto di rapporto di Capacità, chiamato ora indice di Capacità, venne scoperto fuori dal Giappone nel 1974 da un tale Juran che sviluppò l'indice C_p realizzando un bisogno delle aziende, quello di confrontare le specifiche del processo con la variabilità. Il primo indice comunque venne usato per misurare la Capacità di un processo e con il termine Capacità ci si focalizzava sulla frazione di prodotti non conformi alle specifiche. Più avanti però, questo modo di focalizzare solo sulla frazione di prodotti non conformi si spostò verso una Capacità del processo intesa come controllo della qualità di un prodotto solo se vengono controllati i processi produttivi.

L'uso degli indici di Capacità negli Usa comincia quando, negli anni 80, la Ford Motor Company comincia ad usare i C_p e i C_{pk} per misurare le performance dei suoi processi. A quel tempo, l'industria automobilistica venne coinvolta da una grande ventata di miglioramento del proprio sistema di controllo qualità, nei riguardi di se stessi e dei propri fornitori. La pratica di usare gli indici di Capacità quindi si diffuse velocemente in tutte le case produttrici di automobili e intorno ai loro fornitori e questa esplosione coinvolse anche le industrie degli altri settori.

Dal 1991 le tre grandi case automobilistiche statunitensi (General Motor Group, Chrysler Corporation e Ford Motor Company) cominciarono ad usare a pieno regime, per se stessi e per i propri fornitori, il processo di Controllo Qualità e di conseguenza gli Indici di Capacità. Tutte e tre le case automobilistiche, accompagnate dalla società statunitense per il Controllo della Qualità, collaborarono insieme come AIAG (Automotive Industry Action Group (1991)) al progetto di rendere standard sia i manuali di riferimento, sia il formato del reporting e sia la nomenclatura tecnica usata dalla tre compagnie automobilistiche. Il risultato di questo fu l'AIAG Fundamental Statistical Process Control Reference Manual, manuale di riferimento per il Controllo della Qualità di processo, che ufficializza gli indici di Capacità usati e le loro

stime considerate accettabili. Tuttora questo manuale viene considerato come guida al reporting dell'azienda automobilistica e dei suoi fornitori.

La nuova tendenza che portava all'analisi della Capacità del processo continua con molti sforzi, migliorando quindi l'uso del Cp da parte di molte compagnie americane e ottenendo in queste un nuovo approccio alla Qualità attraverso metodi statistici, portando di conseguenza alla riduzione dei costi e al miglioramento della Qualità.

Come può essere notato quindi, gli Indici di Capacità stanno diventando sempre più uno standard per l'informazione riguardante la Qualità di un processo aziendale e dell'azienda stessa, e una corretta spiegazione della loro funzionalità e delle loro proprietà è vitale per diventare e rimanere un fornitore qualificato. Come queste aziende hanno attestato con gli anni, il successo degli Indici di Qualità può essere misurato in dollari risparmiati.

1.3. Struttura di base degli Indici di Capacità e primi problemi relativi al loro uso.

Gli Indici di Capacità sono usati per dare una indicazione sulla Capacità del processo in una forma che diventa facile da capire, rappresentano l'abilità del processo preso in esame di fabbricare prodotti conformi alle specifiche.

Infatti, come dice Zenari, "è vitale conoscere le prestazioni del processo tramite una analisi di capacità dello stesso, lavorando su un campione rappresentativo di osservazioni raccolte, e un modo per quantizzare questa analisi è calcolare un ben determinato indice".

Usare gli Indici di Capacità per esprimere la Capacità di un processo ha reso più semplice raggiungere dei traguardi di Qualità e il loro uso continua a diventare sempre più comune in tutte le aziende. Questo incremento nell'uso è dovuto principalmente al gran numero di aziende che includono nei loro contratti d'acquisto con i fornitori la richiesta di una verifica, per ogni spedizione, della Qualità del loro processo produttivo attraverso gli Indici di Capacità.

Più importante quindi è sottolineare come gli questi Indici indichino al management, grazie ad un singolo numero, un riassunto di cosa stia succedendo all'interno del piano produttivo semplificando così la supervisione delle attività produttive.

L'idea generale che corre accanto agli Indici di Capacità è, quindi, di mettere a confronto quello che il processo "potrebbe fare" con quello che "realmente sta facendo". L'intervallo di specifica riflette i confini dell'utilizzabilità del prodotto secondo i requisiti richiesti dal cliente, controllando così se, dal processo produttivo che abbiamo considerato, otteniamo un prodotto di alta Qualità. Quindi quello che il processo "sta realmente facendo" si riferisce in prima istanza alla variabilità del processo, infatti più bassa è la variabilità, più basso è il rapporto dei prodotti che escono dall'intervallo di specifica. Se il rapporto di Capacità

$$\frac{\text{intervallo di variabilità del processo}}{\text{intervallo di specifica}}$$

è usato per determinare la Capacità del processo, sarebbe meglio fosse quanto più basso possibile. Infatti più basso è il rapporto, più bassa è la dimensione dell'intervallo di specifica usato dai dati del processo.

Quindi un rapporto pari a 1 significa che la variabilità o spread del processo è esattamente uguale all'ampiezza dell'intervallo di specifica, invece un rapporto pari a 0.75 significa che la distribuzione usa il 75% dell'intervallo di specifica.

La forma generale del Rapporto di Capacità ha un senso intuitivo ed è stata molto popolare e usata, ma non si sa perché il suo inverso sia stato usato come forma generale dell'Indice di Capacità di un processo:

$$\frac{\text{intervallo di specifica}}{\text{intervallo di variabilità del processo}}$$

quindi per questa forma generalmente usata la regola fondamentale è: “più grande è il rapporto migliore è”.

Il parametro di paragone per un processo considerato “capace” lo si ottiene quando l’Indice di Capacità vale 1.33, indicando questo una piccolissima frazione di prodotti non conformi alle specifiche.

Gli Indici di Capacità dopo una loro prima definizione, si sono sviluppati con gli anni e hanno raggiunto un numero considerevole soprattutto negli ultimi vent’anni dove si è concentrato il loro sviluppo maggiore e allo stesso momento quindi il loro uso maggiore. Questo loro uso ha portato però, alla nascita di vari problemi, in particolare quello inerente all’opportunità di applicare in modo corretto questi Indici ai casi pratici. Nei primi anni novanta infatti sorge una disputa in cui si sottolinea come ci fosse qualcosa di sbagliato dovuto agli Indici stessi o a chi li usava, i cosiddetti “professionisti” o “pratici”.

Il numero degli Indici di Capacità differentemente studiati e creati da parte dei cosiddetti “statistici” o “teorici” si incrementa sempre più portando quindi alla confusione chi li deve usare, smentendo chi sosteneva fosse stata data una adeguata e chiara spiegazione del significato dei vari Indici. Questo porta ad una stato di confusione e abbandono da parte dei professionisti che quindi sono portati ad usare gli Indici adatti al processo che stanno trattando in modo poco responsabile.

A questo punto, per cercare di correggere questo momento di abbandono e confusione, intervengono due dei più importanti ed autorevoli teorici, Kotz e Johnson, che in vari loro articoli e libri cercano di indirizzare all’uso corretto degli Indici di Capacità chi li deve mettere in pratica.

Loro scopo è quello di chiarire il significato statistico degli Indici di Capacità, presentando e sottolineando la teoria che sta loro dietro, studiando e mettendo a confronto dettagliatamente le proprietà dei vari Indici ed infine cercando di convincere i professionisti che l’opportunità di dare una stima della Capacità di un processo attraverso una paio di valori piuttosto che uno solo, è la cosa migliore. Infatti, secondo loro, “il secondo numero potrebbe essere un indicatore della bontà del primo indicando che ci si è sbagliati nell’accettare solo il primo Indice come un indicatore preciso della Capacità del processo”. Questo quindi indica come si fosse presentata e data, fino a quel momento, maggiore attenzione al cercare di trovare il giusto indice per il processo considerato piuttosto che al cercare l’uso corretto ed appropriato dell’Indice di Capacità.

Recentemente, Kotz e Johnson, per rispondere a questa disputa continua, hanno affermato che “i recenti lavori possono essere considerati come una ricerca fatta per cercare di conciliare i punti di vista dei professionisti e teorici, concentrandosi soprattutto su degli studi dettagliati riguardanti la natura degli Indici di Capacità. Un altro approccio è la determinazione di recenti Indici diventati utilizzabili in un modo più particolare e ricco sebbene comportino alcune volte confusione nel loro uso. Ma i professionisti pensano che

questa varietà porti alla confusione ancora e sempre, ma il nostro studio si concentra su questo approccio e cerca di sottolineare il trend corrente nell'utilizzo degli Indici di Capacità". Sempre a riguardo di questo continuo scontro sulla confusione e incomprensione riguardante l'uso appropriato degli Indici di Capacità nelle aziende, in un loro articolo, Stefan Steiner, Bovas Abrham e Jock Mackay, sottolineano questo problema con un esempio pratico. Infatti, rifacendosi ad un importante manuale di riferimento per la Qualità usato dall'industria automobilistica, il QS9000 standard, manuale che si affida all'uso degli Indici di Capacità, si ricavano i seguenti due estratti. Il primo estratto proviene dalla sezione 4.9.3 del manuale Quality System Requirements (Automotive Industry Action Group(AIAG), 1995) e afferma: "I requisiti sulle performance dei processi in corso sono definite dai clienti. Se questi requisiti non sono stati ottenuti, i seguenti valori di inadempienza appaiono:

- Per un processo stabile e distribuito normalmente si potrebbe ottenere un $Cpk \geq 1.33$
- Per processi non stabili con un output che soddisfa le specifiche e un campione che si può predire si potrebbe ottenere un $Ppk \geq 1.67$ ”.

Il secondo estratto è dal manuale Advanced Product Quality Planning (Automotive Industry Action Group(AIAG), 1995), nel quale i requisiti specifici della Ford Powertrain per un piano di controllo dinamico sono:

“Tutti i processi devono produrre tutte le caratteristiche specifiche di base della produzione e queste caratteristiche devono essere in controllo statistico con $Ppk \geq 1.67$ e $Cpk \geq 1.33$.”

Questi esempi dal QS9000 dimostrano come decisioni importanti riguardanti la soddisfazione delle specifiche richieste del cliente siano basate sul valore che assume l'Indice di Capacità, dimostrando però una certa confusione e incomprensione sul loro esatto utilizzo.

1.4. Il dibattito sull'uso o no degli Indici di Capacità.

Il dibattito presentato precedentemente tra i cosiddetti “professionisti” e “statistici”, ha avuto di sicuro un impatto negativo su quello che è il mondo aziendale e industriale. Infatti, agli occhi degli statistici, gli Indici di per sè appaiono semplici da usare ma per i professionisti invece questi portano ad errate interpretazioni, sostenendo che le loro fondamentali condizioni, come la normalità o la stabilità dei dati, sono troppo rigide e rare da incontrare. Quindi molti milioni di dollari furono spesi nella ricerca sul miglior uso possibile degli Indici di Capacità e, allo stesso tempo, le decisioni che dovevano essere prese in base ad un solo valore provocarono malumori e confusione nei professionisti della Qualità.

Questi ultimi infatti accusarono gli statistici di “terrorismo statistico”, sostenendo come ci fosse un abuso e una manipolazione senza scrupoli degli Indici di Capacità, Indici che cercavano di dare troppe informazioni attraverso un solo ed unico valore difficilmente interpretabile.

Bunker (1991) per esempio, in un suo articolo, definisce terrorismo statistico “l'uso di valide tecniche statistiche attraverso le quali si minaccia e si intimidisce di ottenere un oggetto di business, anche se l'oggetto potrebbe essere ragionevole”. Per capire meglio Bunker fa qualche esempio da lui scoperto:

- Terrorismo del ‘Bandwagon’. I clienti richiedono in modo aggressivo ai propri fornitori di implementare il loro sistema di Controllo Qualità e certe volte gli danno anche un termine massimo per questa loro implementazione, dopo la quale la prova della loro Qualità raggiunta è richiesta in ogni loro spedizione attraverso le Carte di Controllo. Il risultato di questo è che i fornitori ignorano la metodologia statistica del Controllo della Qualità e focalizzano la loro attenzione nel fare Carte di Controllo che possano accontentare i clienti. I fornitori quindi semplicemente non vogliono emettere le vere Carte di Controllo fino a quando la paura di perdere il cliente non sarà svanita, essendo così portati a mentire.
- Terrorismo della ‘verifica delle tasse’. L'uso di standards da parte dei clienti costringono i fornitori alla stima della Capacità secondo le guide linea dei loro clienti, che alcune volte non sono appropriate, per esempio, se i dati sono distribuiti in modo non normale. I standards poco flessibili impediscono ai fornitori di capire i loro processi e di far corrispondere a questi le giuste tecniche di stima per la Capacità. Le conseguenze di questo sono non capire in modo chiaro e corretto i standards dei clienti dando quindi attenzione ai prodotti piuttosto che ai processi, in modo errato dal momento che i processi devono migliorare per produrre prodotti migliori e non viceversa.

I professionisti quindi, come si può capire da questi esempi, vedono questi Indici e questa loro imposizione come il più potente mezzo per misurare la Qualità di un processo produttivo

e li usavano cercando di seguire le condizioni specifiche senza capire quale significato abbiano.

Gli statistici quindi, credenti e creatori degli Indici di Capacità, per cercare di incitare ancora di più l'uso di questi, sostenevano come gli Indici non fossero più difficili da usare di qualsiasi altro mezzo statistico usato per misurare e definire l'output di un processo. Infatti chi usa questi Indici li vede come una variabile determinista piuttosto che casuale e quando chi li usa riuscirà a capire l'uso corretto e costruttivo della variabilità e della distorsione delle stime degli Indici, questi verranno capiti in modo esatto.

Di fatto quindi gli Indici di Capacità portano dei grossi benefici, riassumendo sia la posizione del nostro processo rispetto alla varianza in un solo numero sia riducendo del 50 % i dati che sono necessari per una analisi tradizionale fatta separatamente sulla media e sulla varianza. Benefici però possono venire meno se, come detto precedentemente, viene ad instaurarsi un'atmosfera di terrorismo che va ad infiltrarsi nei processi sull'Analisi della Capacità.

Una volta che gli Indici di Capacità vengono immessi in una azienda o organizzazione che non ha esperienza riguardante la statistica, i clienti cominciano a vedere questi Indici come una medicina per i problemi che loro hanno nel miglioramento della Qualità, cominciando quindi a chiedere settimanalmente o anche giornalmente C_p e C_{pk} per tutti i processi produttivi che ritengono critici, chiedendo rigidamente un requisito minimo per questi due Indici pari a 1, 33. Requisiti che devono essere raggiunti da parte delle aziende fornitrici, dando poca considerazione di conseguenza ai costi di produzione e ottenendo come risultato finale una stima senza senso della Capacità dei processi fuori controllo, usando tecniche che molte volte sono inappropriate. Diventa quindi come un abuso che rischia di trasformarsi in un traguardo irragionevole, che i managers confondono con il reale scopo del controllo statistico della qualità. Fra i maggiori usi delle informazioni tratte dallo studio della Capacità del processo vengono elencati qui i più importanti:

- predire come un processo ben regolato possa rispettare determinate tolleranze
- assistere i tecnologici nelle modifiche del processo
- indicare la frequenza di campionamento per il controllo di processo
- specificare i limiti di validità delle nuove apparecchiature introdotte nel processo
- capacità di selezione tra eventuali fornitori
- pianificare le sequenze del processo produttivo, quando vi è un effetto interattivo del processo sulle tolleranze
- ridurre la variabilità del processo di fabbricazione

Pertanto come si può notare, l'Analisi della Capacità, e quindi di conseguenza gli Indici di Capacità, sono una tecnica che ha applicazioni in molti segmenti del ciclo produttivo includendo il prodotto, la programmazione del processo, la produzione, le quote di vendita e la pianificazione della produzione.

CAPITOLO 2

GLI INDICI DI CAPACITÀ C_p E C_{pk}

2.1. Introduzione.

2.1.1 Sono il C_p e il C_{pk} un ostacolo per il continuo miglioramento?

Esiste una loro alternativa?

In questi ultimi anni sono stati scoperti e usati “nuovi” Indici per migliorare e coprire quelle aree ritenute scoperte dal range di condizioni imposte dagli originali Indici di Capacità giapponesi. Ma nessuno di questi “nuovi” Indici ha mai superato il successo e la notorietà raggiunta dai due Indici di base C_p e C_{pk} . Questi hanno superato il test imposto dal tempo e i cambiamenti imposti dai nuovi concorrenti, rimanendo considerati come gli Indici standard per la quantificazione della Capacità di un processo. C_p e C_{pk} non sono però perfetti come sottolineano molte critiche a loro giunte, ma la loro semplicità e facilità nell’esser capiti li hanno catapultati nel rank delle misure standard della Qualità.

Negli ultimi anni tuttavia, molti articoli del Quality Progress contengono lettere rivolte all’Editore nelle quali si critica (e raramente si difende) l’uso di questi Indici di Capacità, spesso accompagnati da richiami sul loro abbandono dal momento che la loro debolezza era causa di errato uso.

Ciononostante, in un articolo del Quality Progress, il “the Use and Abuse of C_{pk} ” (1989), si sottolinea come importanti documenti statistici che trattano l’appropriato uso dei C_p e C_{pk} siano apparsi solamente nella letteratura tecnica che, la maggior parte delle volte viene tralasciata da chi usa questi Indici, portando a problemi seri ed ad uso scorretto degli Indici di Capacità. Infatti i C_p e C_{pk} non possono essere usati in modo assoluto.

Quindi, perché gli Indici di Capacità di base abbiano un significato dobbiamo considerare un processo sotto controllo, processo che produce dei risultati di stabilità per i parametri misurati. Se questo non fosse rispettato ci indirizzeremmo verso una falsa pista, perché senza il controllo statistico il processo non sarebbe predicabile, quello che faremmo domani dipenderebbe tanto quanto la fase lunare da quello che abbiamo fatto oggi.

Sempre secondo Gunter (Gennaio 1989), il C_{pk} è senza significato perché fornisce solo una fotografia istantanea dalla quale generalizzazioni non possono essere fatte, infatti il C_{pk} di un processo fuori controllo non fornisce informazioni precise su quello che potremmo aspettarci perché niente può essere previsto. Il traguardo da raggiungere sotto queste condizioni, che spesso sono la causa di un uso scorretto degli Indici di Capacità, è quindi quello di identificare ed eliminare le cause responsabili dell'eccessiva variazione ottenendo così la stabilità, e non quello di produrre valori che soddisfano l'arbitrarietà.

Si pretende anche che la distribuzione reale del processo sia conosciuta - qualcosa che quasi mai accade in pratica e che può essere descritta matematicamente - una finzione conveniente quindi e infatti in pratica la miglior cosa che si possa fare è approssimare questa situazione.

Gunter (Giugno 1989) indica anche come gli Indici di Capacità C_p e C_{pk} non possano essere capiti, dal momento che il loro uso richieda l'assunzione di operare con distribuzioni normali. Ma questa assunzione è differente da quelle che noi facciamo quando usiamo statistiche campione - media e deviazione standard - per descrivere e prendere decisioni su ogni tipo di dati?

Queste due statistiche campione sono sempre state usate per descrivere le popolazioni basate su risultati campionari e sono sempre usate anche per accettare e rifiutare decisioni riguardanti la stima della Capacità del processo. Molte poche organizzazioni e aziende possiedono i mezzi necessari per determinare, prima di usare le statistiche, se i dati presi in considerazione siano distribuiti normalmente.

Se noi confrontiamo una curva normale, generata con le statistiche derivanti dai risultati campionari, alla distribuzione della popolazione dalla quale i dati studiati risultano provenire, si può notare delle volte come la deviazione standard non sia capita, dal momento che sovrastima l'ammontare dei dati sui limiti della distribuzione, indicando quindi dei dati fuori dai limiti di specifica.

Quindi, sia Gunter che Paul F. McCoy (Febbraio 1991), sottolineano come poche aziende e poche industrie determinino quanto chiuso rispetto la normalità sia il processo produttivo da loro considerato. Infatti, il numero delle serie aziendali produttive rispecchianti la realtà e distribuite normalmente che portano ad accurate stime della distribuzione della popolazione sono molto probabilmente un'eccezione.

Quindi l'analisi su cosa attualmente occorra e su quale sia la reazione appropriata per questo, non proviene dal riassumere le statistiche, dalle distribuzioni di frequenza o se le

distribuzioni siano normali, ma proviene dalla sequenza temporale dei risultati dei test, che mostrano molto di più di una carta di controllo. Tutto quello che è necessario sono limiti di controllo costruiti per separare, se occorre, la varianza residua da quel segnale statistico che indica qualcosa di indesiderato.

Il fatto dunque è che non c'è una semplice alternativa al C_p o al C_{pk} . Come sottolinea Kearn (Giugno1989): "La qualità non è una bandiera. Non è un modo di parlare. È un lavoro difficile. Il C_p e il C_{pk} sono un tentativo per cercare l'uscita più facile ma non ci sono uscite facili ". Una comprensione esatta della Capacità di un processo, e come si possono superare gli ostacoli trovati, non può riassumersi in un solo numero. Per prendere delle decisioni giuste, le molte situazioni che si verificano devono essere capite, si devono quindi usare carte di controllo, si deve compiere un'analisi statistica appropriata e si devono fare esperimenti progettati. Semplicemente il C_p e il C_{pk} sono gradevoli da usare quando chiariscono e non quando portano a scelte sbagliate. Quindi, quando i concetti di Capacità e controllo non sono capiti da chi prende le decisioni, quando lo studio della Capacità è considerato come qualcosa di forzato piuttosto che come mezzo per il miglioramento della Qualità e quando le conclusioni statistiche cruciali sono ignorate in favore di intuizioni casuali, succede un disastro.

I cosiddetti "professionisti" infatti hanno capito l'importanza che questo disastro assume se si prendono delle decisioni finanziarie troppo rischiose usando stime improprie di C_p e C_{pk} ma hanno anche capito che se si usano stime corrette di C_p e C_{pk} queste portano al successo nel proprio business, dimostrando per concludere che non esiste una loro alternativa essendo necessari e indispensabili per calcolare la Capacità di un processo.

Ad una domanda fatta dai professionisti, "perché non usare μ e σ come misure di Capacità", la risposta dei teorici è "che gli Indici di Capacità hanno il vantaggio, rispetto a queste due statistiche campionarie, di includere un confronto anche con i limiti di specifica del processo, come anche quello di dare una misura senza unita della Capacità che rende i processi in tutto e per tutto comparabili su una scala singola.

Prima quindi di passare alla stima di C_p e C_{pk} bisogna verificare che i dati ricavati dalla caratteristica sotto indagine rispettino determinati vincoli imposti dal particolare tipo di distribuzione a cui gli Indici devono sottostare.

Nella raccolta dell'analisi dei dati per la stima della Capacità di un processo, può succedere che alcune osservazioni si discostino in modo evidente rispetto a tutte le altre. Sorge quindi la necessità di poter individuare, prima di passare alla stima di C_p e C_{pk} , i dati anomali (chiamati anche outliers), grazie a tecniche statistiche e test.

Tecnica abbastanza semplice ed a carattere grafico che permette, sfruttando le proprietà della mediana e delle misure ordinali, di rilevare i potenziali valori sospetti indipendentemente dal tipo di distribuzione e la tecnica denominata "Box and Wisker plot", - letteralmente scatola a baffi. L'uso del "Box and Wisker plot" permette di aumentare

fortemente l'informazione sull'andamento e sulle caratteristiche dei dati di una data variabile presa sotto esame.

Altro importante aspetto per poter usare in modo corretto il C_p e il C_{pk} è che questi due Indici di Capacità di base richiedono che i dati delle caratteristiche messe sotto controllo seguano approssimativamente la distribuzione normale.

Le tecniche che verificano questo aspetto sono il test di Kolmogorov-Smirnov (se la numerosità del campione supera le 50 osservazioni) e il test di Shapiro-Wilk (se la numerosità del campione è minore delle 50 osservazioni).

A questo punto, dopo aver specificato queste tecniche per verificare se le stime che ottengo di C_p e C_{pk} possono aver un significato non portandomi quindi ad un loro abuso, passo allo studio di questi due Indici di Capacità di base.

2.2. L'indice di Capacità C_p .

2.2.1 Definizione di C_p .

L'indice C_p è stato il primo e l'originale dei cinque Indici di Capacità giapponesi, introdotto da Juran nel 1974 ma considerato un Indice accettabile solo dopo gli anni ottanta. È stato calcolato per dare una misura indiretta della potenziale Capacità del processo di soddisfare le richieste dei clienti e non tiene conto se il processo è centrato rispetto ai limiti di specifica. Come sottolineato nella parte precedente, per non causare misure inappropriate della Capacità di un processo e quindi un suo errato uso, l'Indice C_p può essere usato solo sotto ben determinate condizioni:

- a) I dati devono essere distribuiti normalmente: $N(\mu, \sigma^2)$
- b) I dati devono essere incorrelati
- c) Il processo deve essere sotto controllo

A questo punto posso definire il C_p come:

$$C_p = \frac{USL - LSL}{6\sigma}$$

dove **USL** è il limite di specifica superiore mentre **LSL** è il limite di specifica inferiore delle caratteristiche di X considerate e σ è la deviazione standard del processo considerato. I valori di X fuori dai limiti di specifica sono definiti "non conformi" o **NC**.

Chiaramente un valore più alto di C_p è quello che si desidera ottenere, perché un valore basso (soprattutto sotto l'uno) non viene accettato dal momento che indica che il campo di variabilità del processo non va e non potrebbe andar bene con le bande di tolleranza, indicando una incapacità da parte del processo.

Una domanda importante da farsi è: "Ma perché al denominatore si usa come moltiplicatore della deviazione standard il 6?".

Il motivo fondamentale del perché si usi il 6 come moltiplicatore è che, se si considera una distribuzione normale, la probabilità p di ottenere prodotti non conformi (NC) è:

$$p = \Phi\left(\frac{LSL - \mu}{\sigma}\right) + \left\{1 - \Phi\left(\frac{USL - \mu}{\sigma}\right)\right\}$$

dove $\Phi(\bullet)$ è la funzione di ripartizione di una normale standard $N(0, 1)$. Nella figura 1 viene rappresentata una distribuzione normale dove si sottolinea come la probabilità di ottenere prodotti non conformi provenga dai dati che stanno fuori dai limiti di specifica e in particolare per la normale sulle code.

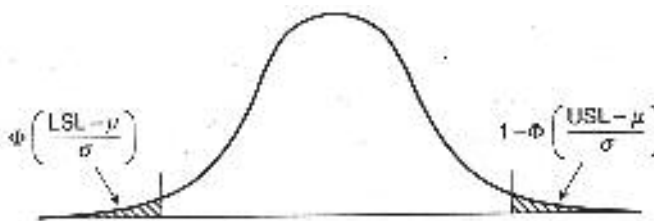


Figura 1

Quindi considerando C_p in termini di vera Capacità del processo con uno sguardo alla percentuale di NC, se la distribuzione è normale e se, il valore atteso di X , è uguale alla quantità $\frac{1}{2} (USL + LSL)$ – punto centrale dell’intervallo di specifica – allora la percentuale di prodotti NC è $2\Phi(-d/ \sigma)$ dove $d = \frac{1}{2} (USL - LSL)$ è la metà della lunghezza dell’intervallo di specifica.

Infatti, dopo aver definito d , posso vedere che

$$C_p = \frac{d}{3\sigma}$$

e così ottenere la percentuale $p\%$ di prodotti non conformi come

$$p\% = 2\Phi(- 3 C_p)$$

Dunque, grazie a questa percentuale di prodotti non conformi, può essere osservato se quello che otteniamo dal processo considerato è la cosa migliore che questo possa fare, dando una certa attenzione al produrre prodotti buoni.

Per chiarire questo, Kocherlakota (1992) sviluppò una tabella che avesse il ruolo di essere una line a guida generale per la percentuale p di prodotti n on conformi, assumendo che il processo fosse perfettamente centrato nel mezzo dell’intervallo di specifica. Questa tabella è:

Campo di variabilità	C_p	Percentuale minima di NC
6σ	1.00	0.0027
8σ	1.33	0.00006334
10σ	1.67	0.0000005733
12σ	2.00	0.000000001973

Quindi se il campo di variabilità è 6σ e il $C_p = 1$, la percentuale di NC stimata è 0.27%, che è vista come una percentuale tutto sommato bassa. Infatti, spesso, per essere accettato è richiesto una valore di $C_p \leq 1.33$.

Ma bisogna però interpretare questa tabella con molta cautela, infatti un $C_p = 1$ non garantisce che ci saranno d’ora in poi solo lo 0.27% di prodotti non conformi ma garantisce

che, con i dati distribuiti in modo normale e perfettamente centrati in $\mu = \frac{1}{2} (USL + LSL)$, non avremmo mai una percentuale di prodotti non conformi inferiore allo 0.27%, producendo quindi un limite inferiore per questa percentuale.

Ora invece, sta cominciando ad avere il sopravvento il trend di esprimere la percentuale di NC in termini di parte per milione, ottenendo quindi la relativa tabella:

Difettosità del processo espresso in parti per milione (ppm)	
C_p	ppm
0.25	453255
0.50	133614
0.60	71861
0.70	35729
0.80	16395
0.90	6934
1.00	2700
1.10	967
1.20	318
1.30	96
1.40	27
1.50	7
1.60	2
1.70	0.34
1.80	0.06
2.00	0.0018

Per illustrare il suo uso possiamo notare come un $C_p = 1$ implichi un tasso di difettosità pari a 2700 parti per milione, notando una certa congruenza con quello che ottenevamo dalla tabella precedente (0.27%) se consideravamo un $C_p = 1$, solamente che in quella si esprimeva la percentuale minima di prodotti non conformi.

Per l'appunto, qualcuno potrebbe stupirsi nel chiedersi perché l'uso della stima della percentuale di NC non sia diventata così popolare quanto il C_p , ma a questo ci hanno pensato Lam e Littig (1992) che hanno cercato di mettere insieme l'idea della percentuale di prodotti non conformi con l'idea del C_p creando così il C_{pp} :

$$C_{pp} = \frac{1}{3} \Phi^{-1} \left(\frac{1}{2} (p + 1) \right)$$

basato sulla stima di p ricavata dai valori osservati della variabile X .

A questo punto ci si chiede: "Ma per quale valore di C_p il nostro processo potrebbe quindi essere considerato capace ?"

Il valore di C_p più comunemente accettato come limite inferiore per la determinazione se un processo sia Capace è $C_p = 1.33$, assicurando dunque che, se il processo è centrato, ci dati del processo stiano più o meno dentro al 75% del campo di variabilità specificato.

Quindi Mizuno, nel 1988, per dare delle linee guida ai manager delle aziende Statunitensi e per dare un uso più dettagliato dei C_p , propose questa tabella:

Indici di Capacità	Valutazione	Risposta
$C_p \geq 1.33$	Capace	E' sufficiente fare delle ispezioni solo all'inizio delle operazioni. Possiamo considerare una accelerazione del processo o altrimenti un aumento di carico
$1 \leq C_p \leq 1.33$	Bisogno di sorveglianza	Pericolo di produrre pezzi difettosi. Bisogno di osservare il processo produttivo considerato
$C_p < 1$	Non capace	Bisogno di cambiare procedure, di cambiare le attrezzature o tararle in modo migliore e di cambiare le tolleranze. Supervisione di tutti gli output del processo.

Ci sono state comunque anche altre considerazioni sul valore corretto del C_p : Montgomery (1996) affermava che il valore minimo da preferire per il C_p fosse 1.33 per un processo già esistente e 1.50 per un processo nuovo, mentre Kotz e Johnson (1993) affermavano che il valore minimo richiesto per un processo fosse 1.50 per un processo già esistente e 1.67 per un processo nuovo.

Ricordiamoci comunque che un grande valore del C_p non assicura sempre l'accettabilità se non sono presenti delle informazioni sulla media del processo.

2.2.2. Caratteristiche di C_p

Come si può notare, l'unico parametro che può essere stimato all'interno della formula del C_p è σ , la deviazione standard di X, la cui stima più naturale è:

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

ottenendo così che la stima di C_p è

$$\hat{C}_p = \frac{USL - LSL}{6s} = \frac{d}{3\sigma} = \frac{\sigma}{s} C_p$$

ed è distribuita come $C_p \frac{\sqrt{n-1}}{\chi_{n-1}^2}$.

Per quanto riguarda il momento r-esimo attorno lo zero, Kotz e Johnson (1993) hanno ricavato che questo è:

$$E[\hat{C}_p^r] = \left(\frac{n-1}{2}\right)^{\frac{1}{2}r} \frac{\Gamma(\frac{1}{2}(n-1-r))}{\Gamma(\frac{1}{2}(n-1))}$$

dove ponendo $r = 1$ e $r = 2$ si ottiene rispettivamente il momento primo, $E = E[\hat{C}_p]$ e il momento secondo $SD = \sigma(\hat{C}_p/C_p)$

A questo punto, dal momento che \hat{C}_p come le altre statistiche è condizionata dalla varianza campionaria, è importante trovare l'intervallo di confidenza che includa il vero valore di C_p con un'alta probabilità.

Si comincia per prima cosa notando che

$$\hat{C}_p \rightarrow \frac{C_p}{\sqrt{\frac{X_{n-1}^2}{n-1}}}$$

successivamente osservando che l'intervallo di confidenza $100(1-\alpha)\%$ per σ^2 è

$$\left(\frac{(n-1)s^2}{X_{n-1,1-\alpha/2}^2}, \frac{(n-1)s^2}{X_{n-1,\alpha/2}^2} \right)$$

e giungendo così alla conclusione che l'intervallo di confidenza $100(1-\alpha)\%$ per C_p è pari a:

$$\left(\frac{X_{n-1,\alpha/2}^2}{\sqrt{n-1}} \cdot \frac{USL - LSL}{6\sigma}, \frac{X_{n-1,1-\alpha/2}^2}{\sqrt{n-1}} \cdot \frac{USL - LSL}{6\sigma} \right)$$

o, dal momento che $\hat{C}_p = \frac{USL - LSL}{6s}$, a:

$$\left(\frac{X_{n-1,\alpha/2}^2}{\sqrt{n-1}} \hat{C}_p, \frac{X_{n-1,1-\alpha/2}^2}{\sqrt{n-1}} \hat{C}_p \right)$$

dove $X_{n-1,1-\alpha/2}^2$ e $X_{n-1,\alpha/2}^2$ sono la percentuale dei punti ad $\alpha/2$ inferiore ed $\alpha/2$ superiore della distribuzione Chi-Quadrato con $v = n - 1$ gradi di libertà. Questi valori sono tabulati all'interno della tavola dei valori dei percentili della distribuzione χ^2 .

L'accuratezza della stima del para metro dipende pesantemente dalla grandezza del campione considerato, dalla tecnica di campionamento e da altri fattori condizionanti. Un mezzo per qualificare se la stima di una statistica è buona oppure no, proviene dall'intervallo di confidenza, intervallo che, piuttosto che dare una stima singola del parametro, fornisce una serie di valori all'interno dei quali la nostra stima viene a trovarsi con sicurezza.

Passando invece dall'intervallo di confidenza alla **verifica d'ipotesi**, si osserva come quest'ultima determini se il valore ipotizzato di un parametro sia vero o no, dando quindi maggiori informazioni sul vero valore.

Per quanto riguarda quindi la verifica d'ipotesi su C_p , è quindi necessario dimostrare come questo sia uguale o maggiore di un valore obiettivo c , comunemente fissato a 1.33, giungendo quindi alle seguenti due ipotesi:

$$H_0 : \text{il processo non è Capace se } C_p < c$$

$$H_1 : \text{il processo è Capace se } C_p \geq c$$

Da questo test si ottiene poi il cosiddetto **p-value**, valore che si riferisce al rischio di concludere in modo scorretto che il processo considerato sia capace dato un valore obiettivo e tipicamente se il p-value è maggiore di 0.05, vale a dire il 5%, posso accettare l'ipotesi nulla e dimostrare che il processo non è Capace.

Hoffman (1993), riuscì anche a trovare un metodo generalizzato per il test d'ipotesi su \hat{C}_p , metodo che si basava nel definire $C_p(\text{high})$ come una Capacità del processo che si può ben accettare con una probabilità pari a $p = (1-\alpha)$, nel definire $C_p(\text{low})$ come una Capacità del processo che deve essere respinta con una probabilità $p = (1-\beta)$ e quindi infine nel sviluppare la tabella sottostante che definisce i valori di $C_p(\text{high})/C_p(\text{low})$ e $c/C_p(\text{low})$ per varie numerosità campionarie n e per $\alpha=\beta=0, 10$ o $\alpha=\beta=0, 05$.

Numerosità Campionaria	$\alpha = \beta = 0, 10$		$\alpha = \beta = 0, 05$	
	$C_p(\text{high})/C_p(\text{low})$	$c/C_p(\text{low})$	$C_p(\text{high})/C_p(\text{low})$	$c/C_p(\text{low})$
10	1.8127	1.3601	2.1555	1.5113
20	1.5111	1.2357	1.7014	1.3233
30	1.3979	1.1865	1.5385	1.2523
40	1.3354	1.1587	1.4503	1.2129
50	1.2946	1.1402	1.3935	1.1872
60	1.2655	1.1259	1.3532	1.1688
70	1.2433	1.1167	1.3228	1.1548
80	1.2258	1.1086	1.2988	1.1437
90	1.2115	1.1020	1.2794	1.1347
100	1.1995	1.0964	1.2632	1.1271

Quindi la determinazione di un valore critico appropriato richiede di stabilire sia un livello di qualità di accettazione (LQA) sia un livello di qualità di rifiuto (LQR) e in termini di Capacità di un processo bisogna accettare un processo che ha C_p maggiore di $C_p(\text{high})$ e rifiutare un processo ha un C_p minore di $C_p(\text{low})$.

Per concludere quindi, come visto l'Indice C_p giudica se un processo è Capace attraverso il confronto tra la variabilità del processo stesso e i limiti di specifica, potendo di conseguenza sottolineare come la sua maggior debolezza sia quella di non considerare l'eventualità della posizione della media μ . Quindi C_p non da indicazioni sulle reali performance del processo, non riflette l'impatto che i salti dalla media hanno sulla Capacità del processo di produrre prodotti all'interno delle specifiche.

A questo punto è importante definire quindi l'Indice C_{pk} , che oltre a riflettere i cambiamenti nella varianza del processo indica anche il raggiungimento del valore obiettivo T, che in questo caso è la media, misurando quindi la Capacità attuale e tenendo conto della sua centratura.

2.3. L'Indice di Capacità C_{pk} .

2.3.1. Definizione di C_{pk} .

L'indice C_{pk} è stato creato, come detto precedentemente, dai Giapponesi per sopperire alle lacune dell'Indice C_p , lacune che erano dovute principalmente alla Capacità definita rispetto al valore del 6σ e non tenendo quindi conto della media del processo. Usando solamente il C_p , le aziende Giapponesi che avevano all'interno processi con bassa variabilità non sentivano la necessità di centrare i loro processi produttivi all'interno dei limiti di specifica dati dagli ingegneri, cercando quindi solamente di abbassare la variabilità del processo pensando di risparmiare così soldi sulle materie prime. Ma questo modo di pensare portò quindi in prima istanza ad una perdita nelle performance dei prodotti e quindi successivamente, notate queste perdite, alla definizione del nuovo Indice C_{pk} .

Prima di passare alla sua definizione, ricordiamo che, per non causare misure inappropriate della Capacità di un processo e quindi un suo errato uso, l'Indice C_{pk} può essere usato solo sotto ben determinate condizioni come C_p :

- a) I dati devono essere distribuiti normalmente: $N(\mu, \sigma^2)$
- b) I dati devono essere incorrelati
- c) Il processo deve essere sotto controllo

Prima di definire l'Indice C_{pk} vengono definiti

$$C_{pu} = \frac{USL - \mu}{3\sigma} \text{ e } C_{pl} = \frac{\mu - LSL}{3\sigma}$$

e quindi successivamente

$$C_{pk} = \min(C_{pu}, C_{pl}) = \frac{d - |\mu - m|}{3\sigma}$$

dove il valore $d = (USL - LSL)/2$ è il punto centrale della lunghezza intervallo di specifica mentre il valore $m = (USL + LSL)/2$ è il punto centrale dell'intervallo di specifica.

Il denominatore 3σ sta ad indicare quanto spazio c'è bisogno a destra e a sinistra della media per contenere la maggior parte dei dati, quindi misura quanto variabile sia il processo. Quando la distribuzione è simmetrica si ha la stessa quantità di spazio mentre quando non è simmetrica si alcuni problemi, ed è per questo che quindi si cerca di applicare l'Indice C_{pk} a distribuzioni simmetriche, come per esempio la Normale.

Il numeratore $d - |\mu - m|$ invece indica la quantità di spazio esistente tra la media μ del processo e i limiti di specifica, essendo quindi una misura di quanto bene il processo sia centrato.

A questo punto possiamo notare come

$$C_{pk} \leq C_{pu} \text{ implichi } USL \geq \mu + 3 \cdot \sigma \cdot C_{pk} \text{ e } C_{pk} \geq C_{pl} \text{ implichi } LSL \leq \mu + 3 \cdot \sigma \cdot C_{pk}$$

Dall'osservazione della definizione del C_{pk} appare evidente come questo Indice determini la Capacità della metà peggiore dei dati, quella con minore Capacità verso uno dei due limiti di specifica. C_{pu} e C_{pl} misurano quanto chiuse siano le due code estreme della distribuzione del processo rispetto i limiti di specifica, infatti la varianza del processo potrebbe oscillare intorno a dei valori ma fino a quando la posizione delle code estreme della distribuzione del processo rimangono nello stesso posto, il C_{pk} non cambia. Tutto questo può essere ben sottolineato dalla seguente figura2 in cui si mostrano tre distribuzioni con varianze diverse ma tutte con $C_{pk} = 1$, potendo quindi concludere che, presi singolarmente gli effetti della posizione del processo e la sua variabilità nel calcolo della Capacità del processo, questi vengono confusi all'interno di questo Indice.

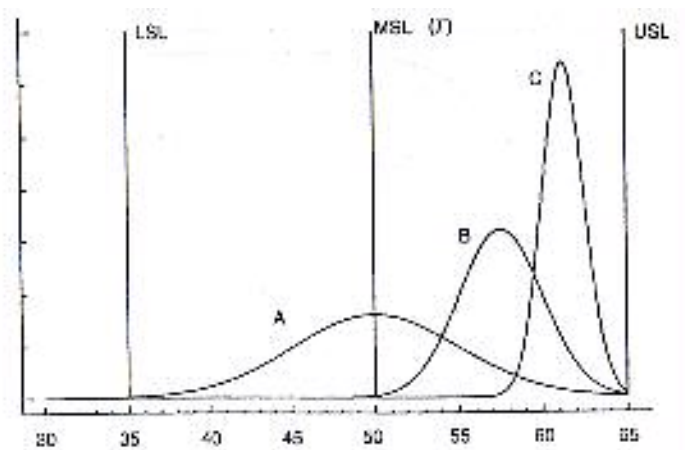


Figura 2

A questo punto anche per l'Indice C_{pk} , come per C_p , può essere calcolata la probabilità p di ottenere prodotti non conformi (NC), infatti sempre sotto la condizione di normalità dei dati abbiamo che questa probabilità è pari a:

$$p = \Phi\left(\frac{LSL - \mu}{\sigma}\right) + \left\{1 - \Phi\left(\frac{USL - \mu}{\sigma}\right)\right\}$$

e la percentuale di prodotti non conformi la si ottiene moltiplicando p per 100, ricordando sempre che $\Phi(\bullet)$ è la funzione di ripartizione di una normale standard $N(0, 1)$.

Kotz e Johnson (1995) hanno calcolato anche i limiti superiori ed inferiori di p , che risultano pari a:

$$\Phi(-3C_{pk}) \leq p \leq 2 \cdot \Phi(-3C_{pk})$$

L'esatta percentuale di prodotti non conformi può essere espressa in funzione di C_{pk} e C_p come:

$$p\% = 100 \cdot [\Phi(-3(2C_p - C_{pk})) + \Phi(-3C_{pk})]$$

Successivamente Castagliola, rifacendosi alla definizione di p , ha dimostrato che la stima di C_{pk} può essere data da:

$$\hat{C}_{pk} = \min \left\{ \frac{\bar{x} - LSL}{3s}, \frac{USL - \bar{x}}{3s} \right\}$$

e scritta come

$$C_{pk} = \frac{1}{3} \min \left\{ -\Phi^{-1} \left(\frac{LSL - \bar{x}}{s} \right), \Phi^{-1} \left(\frac{\bar{x} - USL}{s} \right) \right\}$$

Dalla percentuale di prodotti non conformi può anche essere ottenuto il rendimento percentuale $r\%$ del processo, che è pari a 100 meno la percentuale di prodotti non conformi:

$$r\% = 100 \cdot \left[\Phi \left(\frac{USL - \mu}{\sigma} \right) - \Phi \left(\frac{LSL - \mu}{\sigma} \right) \right]$$

e i suoi limiti inferiore e superiore sono:

$$100 \cdot [2\Phi(3C_{pk}) - 1] \leq r\% \leq 100 \cdot [\Phi(3C_{pk})]$$

Successivamente alla definizione del rendimento r di un processo, Wierda (1994) sviluppò una approssimazione del C_{pk} in funzione del rendimento del processo che può essere applicata abbastanza bene in molte situazioni. Definendo con θ = rendimento del processo, si ottenne

$$\tilde{C}_{pk} \cong \frac{1}{3} \Phi^{-1}(\theta)$$

Questa è una approssimazione cauta particolarmente per i processi centrati all'interno dell'intervallo di specifica e lavora abbastanza bene per processi sufficientemente capaci, per esempio con $C_{pk} > 1$ come mostra la figura3, dove si sottolinea la relazione tra il vero C_{pk} (la curva triangolare) e l'approssimazione del C_{pk} (la curva rotondeggiante) .

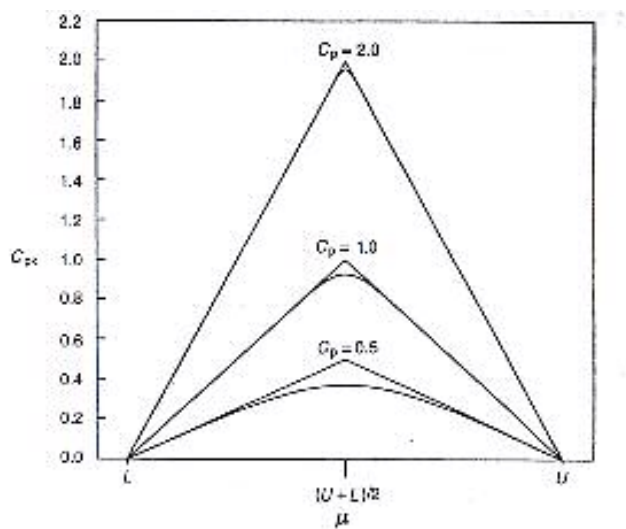


Figura 3

Venne quindi data da Finley (1992) la seguente tabella, usata come linea guida, nella quale si sottolineava la percentuale di prodotti non conformi per un dato valore dell'Indice C_{pk} . Indice che non poteva essere trasformato in una stima specifica del rendimento del processo ma solamente in una stima approssimativa nella forma di limiti superiori ed inferiori.

C_{pk}	Parti al di fuori dell'intervallo di specifica
0.25	16 di 100 sono fuori
1	7 di 100 sono fuori
1.33	13 di 10.000 sono fuori
1.67	1 di 1.000.000 sono fuori
2.00	1 di 1.000.000.000 sono fuori

Per determinare la Capacità di un processo l'Indice C_{pk} è uno dei più usati, e il suo valore di riferimento grazie al quale si può affermare che il nostro processo sia Capace è ancora 1.33, lo stesso del C_p . Infatti un valore di C_{pk} pari a 1.33 sta ad indicare come il processo considerato usi il 75% del campo di variabilità, a differenza del C_p che per un valore pari a 1.33 indica come il processo produca all'interno dell'intervallo di specifica e come usi solamente il 75% del campo di variabilità se il processo è perfettamente centrato nel punto centrale dell'intervallo di specifica. Quindi C_p è un limite superiore per C_{pk} e quando il processo è perfettamente centrato si ottiene che $C_p = C_{pk}$.

Quindi l'Indice C_{pk} mette a confronto la posizione delle code più esterne della distribuzione ai rispettivi limiti di specifica, usando poi il più piccolo dei due confronti come misura della Capacità del processo.

Successivamente, esaminando le relazioni tra gli Indici C_p e C_{pk} , e usando questi due come una coppia, venne scoperto che una tremenda quantità di informazioni potevano essere ricavate sul processo grazie all'Indice di Capacità k , definito come

$$k = \frac{|\bar{M} - \mu|}{1/2 \cdot (USL - LSL)} = \frac{|\mu - 1/2(LSL - USL)|}{1/2 \cdot (USL - LSL)}$$

Questo Indice descrive la Capacità semplicemente come una misura scalare della mancanza di centratura del processo e non può essere comparato con il C_p o il C_{pk} . Quando $k = 0$ il processo è centrato nel valore previsto, quando $k = 1$ la media del processo è posizionata in uno dei due limiti di specifica e quando $0 < k < 1$ la media è posizionata tra il valore previsto e uno dei due limiti di specifica.

Una stima dell'Indice k è:

$$\hat{k} = \frac{|\bar{x} - 1/2(LSL - USL)|}{1/2 \cdot (USL - LSL)}$$

ed è relazionato a C_p e C_{pk} dalla relazione

$$C_{pk} = (1 - k) \cdot C_p$$

Questa riduzione ad un solo Indice può, però, essere pagata con il dimenticarsi le informazioni che abbiamo separatamente sulla media μ e sulla dispersione σ . Infatti nessuno di questi indici racconta la storia vera, dal momento che C_p e k misurano la Capacità rispettivamente in termini di variazione e centratura del processo

2.3.2 Caratteristiche di C_{pk}

L'indice C_{pk} contiene, come il C_p , solo due parametri da stimare, μ e σ . Prima di cominciare però, per avere un uso corretto della stima del C_{pk} , bisogna assicurarsi grazie ad uno studio accurato della Capacità del processo che i dati siano distribuiti normalmente, siano indipendenti tra di loro e siano in controllo statistico.

Come per C_p quindi, una stima di σ e una di μ sono:

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

Una stima naturale di C_{pk} diventa quindi:

$$\hat{C}_{pk} = \frac{d - \left| \bar{x} - \frac{1}{2}(LSL + USL) \right|}{3s} = \min(\hat{C}_{pu}, \hat{C}_{pl})$$

dove

$$\hat{C}_{pu} = \frac{USL - \bar{x}}{3s} \quad \text{e} \quad \hat{C}_{pl} = \frac{\bar{x} - LSL}{3s}$$

A questo punto anche per \hat{C}_{pk} , come per \hat{C}_p , si ricava che il suo momento r-esimo pari a

$$E[\hat{C}_{pk}^r] = \left(\frac{d\sqrt{n-1}}{3\sigma} \right) \cdot r \cdot E[X_{n-1-r}] \cdot \sum_{j=0}^r (-1)^j \cdot \binom{r}{j} \cdot \left(\frac{\sigma}{d\sqrt{n}} \right)^j \cdot E \left[\left| \frac{\sqrt{n} \left[\bar{X} - 1/2(USL - LSL) \right]}{\sigma} \right| \right]$$

e in particolare ponendo $r = 1$ e $r = 2$ mi ricavo rispettivamente il momento primo $E = E[\hat{C}_{pk}]$ e il momento secondo $SD = \sigma(\hat{C}_{pk}/C_{pk})$.

Lo sviluppo del calcolo degli **intervalli di confidenza** per \hat{C}_{pk} non è stato molto facile dal momento che ha caratteristiche distribuzionali più complesse rispetto al \hat{C}_p . Infatti la distribuzione di questo Indice di Capacità coinvolge la distribuzione di due variabili casuali distribuite secondo una t di Student non centrata.

Heavlin (1988) diede questa definizione dell'intervallo di confidenza $100(1-\alpha)\%$:

$$\text{Limite inferiore: } \hat{C}_{pk} - z_{1-\alpha/2} \left\{ \frac{n-1}{9n(n-3)} + \hat{C}_{pk}^2 \frac{1}{2(n-3)} \left(1 + \frac{6}{n-1} \right) \right\}^{1/2}$$

$$\text{Limite superiore: } \hat{C}_{pk} + z_{1-\alpha/2} \left\{ \frac{n-1}{9n(n-3)} + \hat{C}_{pk}^2 \frac{1}{2(n-3)} \left(1 + \frac{6}{n-1} \right) \right\}^{1/2}$$

dove $\Phi(z_\alpha) = \alpha$.

Successivamente Wasserman e Franklin (1992) modificarono la dicitura dei limiti inferiore e superiore dell'intervallo di confidenza, ottenendo così la stima più appropriata e più usata per l'intervallo di confidenza $100(1-\alpha)\%$ per $n \geq 30$ pari a

$$\hat{C}_{pk} - z_{1-\alpha/2} \left(\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)} \right)^{1/2}, \quad \hat{C}_{pk} + z_{1-\alpha/2} \left(\frac{1}{9n} + \frac{\hat{C}_{pk}^2}{2(n-1)} \right)^{1/2}$$

Ci sono anche altre formule, soprattutto per il limite inferiore $100(1-\alpha)\%$, come quella di Dovich (1992) nella quale si esprime l'intervallo di confidenza $100(1-\alpha)\%$ come

$$\hat{C}_{pk} \left[\frac{1 - z_{1-\alpha/2}}{\sqrt{2n-2}} \right], \quad \hat{C}_{pk} \left[\frac{1 + z_{1-\alpha/2}}{\sqrt{2n-2}} \right]$$

Come può essere dedotto quindi, C_{pk} come C_p è Indice di cui non si conosce mai il vero valore con certezza e che quindi deve essere stimato. La stessa teoria che è stata applicata nel fare la **verifica d'ipotesi** per C_p , può essere applicata anche per C_{pk} . Un utilizzo appropriato nel qualificare la stima è però più importante nel caso del C_{pk} , dal momento che questo Indice di Capacità è comunemente più usato nelle decisioni prese da parte dei manager. La qualificazione della stima, se è corretta o errata, proprio come detto per C_p , avviene grazie agli intervalli di confidenza o attraverso la verifica di ipotesi. La verifica d'ipotesi si svolge nello stesso modo di quella del C_p , quindi in una analisi dell'Indice di Capacità del processo C_{pk} , nella quale è necessario dimostrare che C_{pk} sia uguale o maggiore di un valore obiettivo detto C_{p0} .

Abbiamo quindi le seguenti due ipotesi:

$$H_0 : \text{il processo non è Capace se } C_{pk} < C_{p0}$$

$$H_1 : \text{il processo è Capace se } C_{pk} \geq C_{p0}$$

Anche in questo caso, come per C_p , il valore più comunemente usato per il valore obiettivo C_{p0} è 1.33 o 1.67 e il metodo più usato per verificare queste ipotesi è sempre quello di Hoffman (1993). Metodo che si basava nel definire C_p (high), come la Capacità del processo

che si può ben accettare con una probabilità pari a $p = (1-\alpha)$, nel definire $C_p(\text{low})$ come la Capacità del processo che deve essere respinta con una probabilità $p = (1-\alpha)$, e quindi infine nel sviluppare la tabella (riportata all'interno del paragrafo 2.4) che definisce i valori di $C_p(\text{high})/C_p(\text{low})$ e $c/C_p(\text{low})$ per varie numerosità campionarie n e per $\alpha = \beta = 0,05$ o $\alpha = \beta = 0,10$.

Per concludere, anche il C_{pk} , come il C_p , è un Indice di Capacità abbastanza usato per la certificazione dei fornitori ma la sua fragilità ha convinto molti autori a fermare il suo uso. Infatti, una delle maggiori critiche fatte al C_{pk} , è quella delle condizioni stringenti sotto le quali deve essere applicato per assumere un significato, in particolare la normalità dei dati, l'indipendenza fra questi e il processo che deve essere in controllo statistico.

Chou (1990) ha trovato anche che, oltre ad essere il processo in controllo statistico, molte numerosità campionarie hanno bisogno di conseguire una precisione e confidenza accettabile. Per esempio, per un campione di numerosità campionaria pari a 275 (campione di numerosità campionaria difficilmente ottenibile in pratica), ha trovato che è ragionevole ottenere con certezza un C_{pk} reale che è inferiore del 10% rispetto al C_{pk} stimato.

Boyles (1991) sottolinea come "il C_{pk} effettivamente non misuri l'effetto della centratura sul valore obiettivo del processo, non facendo una chiara distinzione tra il processo è sulla strada giusta o il processo è fuori strada, ma sia essenzialmente una misura del rendimento del processo". Più importante ancora, il C_{pk} da solo non dà nessuna indicazione sulla direzione verso la quale il processo sia fuori strada, ma sia la posizione sia la variabilità del processo sono usati solamente per determinare la Capacità. Infatti secondo Mirabella (1991) il C_{pk} , come altri Indici, ha il problema di cercare di misurare due effetti (come per esempio la centratura e la dispersione) in un solo indice, confondendo quindi questi effetti.

Come visto dalla nostra analisi, il vero valore del C_{pk} difficilmente lo si conosce perché quasi mai si conosce il vero valore di μ e σ , valori che quindi vanno stimati. Ci sono due metodi di stimare σ , che dipendono entrambi dalla numerosità campionaria del campione preso in considerazione.

Questo porta Gunter (Maggio 1993) a fare una domanda: "Daranno queste due procedure lo stesso risultato per C_{pk} ? Se no qual'è corretto?". La risposta data è quella che: "A meno che il processo non sia in controllo statistico, avremmo delle risposte differenti. Ed inoltre queste dipendendo da cosa vogliamo fare con le informazioni sul C_{pk} , comunque entrambe le risposte sono sbagliate se il nostro processo non è in controllo statistico".

Quindi come sottolineato anche nella parte iniziale, se non sono soddisfatte tutte le condizioni per poter usare in modo corretto l'Indice di Capacità C_{pk} , avremmo un risultato la cui interpretazione non sarà esatta e porterà fuori strada.

CAPITOLO 3

L'INDICE C_{pm} E ALTRI SUOI INDICI ASSOCIATI

3.1. Introduzione.

Come visto nel capitolo precedente, sia l'Indice C_p che C_{pk} sono usati per sottolineare i cambiamenti nell'ammontare di prodotti oltre i limiti di specifica. Infatti l'Indice C_p giudica se un processo è Capace comparando la variabilità del processo stesso con i limiti di specifica mentre C_{pk} mette a confronto la posizione delle code più esterne della distribuzione con i rispettivi limiti di specifica, usando poi il più piccolo dei due confronti come misura della Capacità del processo. Quindi nessuno dei due considera il valore obiettivo da raggiungere per il processo considerato, processo che è più critico nei confronti delle performance del processo piuttosto che sul range di specifica costruito attorno ad esso.

Gli Indici di base considerati – C_p , C_{pk} , C_{pl} , C_{pu} e k – incorporano infatti l'approccio tradizionale alla qualità, che considera tutti i prodotti all'interno delle specifiche come "buoni" mentre tutti i pezzi fuori dalle specifiche come "cattivi", incurante della vicinanza al valore obiettivo.

La riduzione della variabilità sta molto a cuore per il continuo miglioramento della qualità e, mentre l'approccio vecchio al controllo della qualità guardava solamente al campo di variabilità, ora bisogna concentrarsi non su cosa sia minimamente accettabile ma su cosa assicuri la massima performance.

Proprio per questo vengono introdotti e pubblicati da parte di Chan (1988), e successivamente Boyles (1991), l'Indice C_{pm} e i suoi associati, definendo come "questi Indici siano la misura appropriata del progresso verso i paradigmi del miglioramento della qualità,

paradigmi che hanno la riduzione della variabilità come regola guida e il rendimento del processo come misura primaria del successo”.

In seguito all'Incontro Annuale dell'Associazione degli Statistici Americani presso Las Vegas, viene proposto da Hsiang e Taguchi (1985) un uso particolare dell'Indice C_{pm} , quello che impiega l'approccio della **funzione di perdita** e in cui ci si focalizza sulla riduzione della variabilità attorno al valore obiettivo. Questo approccio alla qualità è differente rispetto l'approccio tradizionale in cui non si aveva nessuna perdita della qualità se il processo non avesse rispettato i limiti di specifica superiore ed inferiore.

Taguchi infatti usa la funzione di perdita quadratica per descrivere una nuova idea, idea che sottolinea come un processo non riveli nessuna perdita solamente se produce e lavora secondo il suo valore obiettivo, sottolineando come una piccola deviazione dal valore obiettivo sia causa di perdita di qualità.

Quindi, prima di definire l'Indice C_{pm} , è necessario fare una precisazione sulla funzione di perdita. C_{pm} e i suoi Indici associati considerano la varianza come $\sigma^2 = E(X-\mu)^2$, dove X è una variabile casuale che descrive la variabile osservata. Questa definizione della varianza determina l'ammontare della variabilità attorno alla media del processo, media che è una informazione necessaria per ottenere la Capacità, ma non fornisce informazioni riguardanti la variabilità attorno al valore obiettivo. Una misura migliore della varianza è quindi $\tau^2 = E(X-T)^2$ dove τ è l'errore quadratico medio dal valore obiettivo.

A questo punto, Hsiang e Taguchi (1985), hanno notato come ogni valore della caratteristica X considerata procuri della perdite monetarie a chi opera lontano dal valore obiettivo. Queste perdite monetarie vengono definite come $L(x) = k(x-T)^2$, e nessuna perdita avviene se la caratteristica X è perfettamente uguale a T. Così, con questo nuovo approccio rivolto al miglioramento della qualità, può essere distinta la differenza tra le caratteristiche misurate racchiuse all'interno dei limiti di specifica e quelle verso il valore obiettivo, potendo quindi osservare come la funzione di perdita di Taguchi modelli una perdita che è una funzione quadratica della distanza dal valore obiettivo. (figura1)

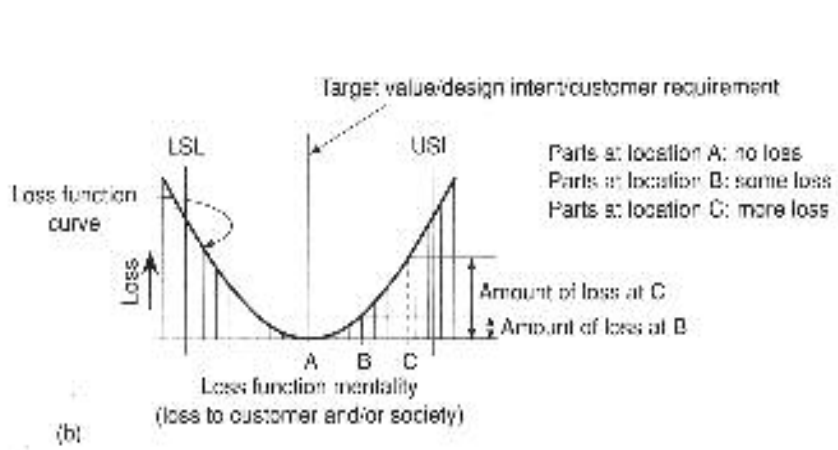


Figura 1

Questo nuovo approccio assume quindi che la funzione di perdita sia proporzionale a $(X - T)^2$ così che $\sigma^2 + (\mu - T)^2$ diventi la misura della perdita attesa. Mirabella e Spring (1991) affermano che una funzione di perdita ragionevole sarebbe costante fuori dai limiti di specifica, probabilmente in questa forma:

$$k(X - T)^2 \quad \text{per } LSL \leq X \leq USL$$

$$k(LSL - T)^2 \quad \text{per } X \leq LSL$$

$$k(USL - T)^2 \quad \text{per } X \geq USL$$

Boyles (1991) notò anche come la Capacità di un processo possa essere espressa dalla funzione di perdita attesa, definita come

$$E(L) = k \cdot E\{(X - T)^2\} = k \cdot \tau^2$$

Questa definizione contiene il concetto di variazione attorno al valore obiettivo, espressa da τ , che può essere espressa anche nella forma alternativa di $\tau^2 = \sigma^2 + (\mu - T)^2$ dove si sottolinea la varianza attorno alla media e la varianza della media del processo attorno al valore obiettivo.

3.2. L'Indice di Capacità C_{pm} .

3.2.1. Definizione di C_{pm} .

Quindi, dopo questa introduzione, può essere definito l'indice C_{pm} come

$$C_{pm} = \frac{d}{3\sqrt{\sigma^2 + (\mu - T)^2}}$$

dove μ è il valore atteso e σ la deviazione standard della caratteristica considerata X , T è il valore obiettivo e $d = \frac{1}{2}(\text{USL} - \text{LSL})$. Il valore obiettivo corrisponde a $m = \frac{1}{2}(\text{USL} + \text{LSL})$, il punto centrale dell'intervallo di specifica, ma se questo caso non si verifica abbiamo dei seri problemi nell'usare il C_{pm} .

Lo sviluppo di questo indice è quindi fatto sotto due condizioni base:

- il processo considerato è distribuito normalmente
- il valore obiettivo T si trova alla metà tra i limiti di specifica in $m = \frac{1}{2}(\text{USL} + \text{LSL})$

Dalla definizione data è quindi possibile notare come C_{pm} possieda le necessarie proprietà richieste per essere considerato un Indice di Capacità. Si vede infatti che se la varianza del processo aumenta (o diminuisce), aumenta (o diminuisce) il denominatore e quindi l'Indice diminuisce (o aumenta), come pure se la media del processo si sposta lontano (o si avvicina) dal valore obiettivo il denominatore aumenta (o diminuisce) e quindi C_{pm} diminuisce (o aumenta).

Come si può notare, il denominatore di C_{pm} è tre volte la deviazione quadratica media di X dal valore obiettivo T

$$3\sqrt{\sigma^2 + (X - T)^2} = 3\sqrt{E(X - T)^2}$$

ed è quindi chiaro che

$$C_{pm} \leq C_p \text{ e } C_{pm} < \frac{d}{3 \cdot |\mu - T|}$$

dal momento che

$$\sigma^2 + (\mu - T)^2 \geq \sigma^2$$

Parlar e Wesolowsky riuscirono a trovare, partendo dalla definizione di C_{pm} e riprendendo quella di C_p , che tra questi due indici esisteva una relazione pari a

$$C_{pm} = \frac{C_p}{\sqrt{1 + 9(C_p - C_{pk})^2}} = \frac{C_p}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

e successivamente anche una relazione tra il C_{pm} e il C_{pk} pari a

$$C_{pk} = C_p - \frac{1}{3} \sqrt{\left(\frac{C_p}{C_{pm}}\right)^2 - 1}$$

trovando quindi che

$$C_p \geq \max(C_{pm}, C_{pk})$$

Si scoprì a questo punto che per molti pratici era difficile interpretare cosa significasse veramente un dato valore di C_{pm} , osservando come per massimizzare questo Indice si dovesse minimizzare la quantità $\sigma^2 + (\mu - T)^2$. Il C_{pm} quindi, potrebbe essere incrementato sia dal ridurre la variabilità del processo sia dall'aver la media del processo chiusa rispetto al valore obiettivo, non essendo però ancora chiaro quale di questi due approcci ottimizzi il processo.

Kotz e Johnson (1993) hanno anche cercato di trovare, riuscendoci, delle relazioni che legassero il C_p , il C_{pk} e il C_{pm} sotto varie condizioni.

In prima istanza sono giunti alla conclusione che

- se si verifica $T = \mu$, allora il C_p è uguale al C_{pm}
- se si verifica $\mu = m$, allora il C_{pk} è uguale al C_{pm} .

Osservando successivamente se la caratteristica del processo ha una distribuzione normale, Kotz e Johnson hanno concluso che, un valore pari a uno per tutti e tre gli Indici, C_p , C_{pk} e C_{pm} , significa una percentuale attesa del prodotto con valore di X all'interno dei limiti di specifica inferiore al 99.7%. Valori negativi invece sono impossibili sia per C_p che per C_{pm} .

La principale distinzione tra il C_{pk} e il C_{pm} è, come sottolinea Mirabella (1991), quella di dare una importanza diversa ai limiti di specifica USL e LSL rispetto al valore obiettivo T . Infatti la funzione principale del C_{pk} è quella di indicare la Capacità del nostro processo tenendo in considerazione la sua media, quindi per $LSL < \mu < USL$ il $C_{pk} \rightarrow \infty$ se $\sigma \rightarrow 0$, notando però come un grande valore di questo Indice non procuri informazioni sulla discrepanza tra μ e T . L'Indice C_{pm} invece misura il grado di raggiungimento del valore obiettivo da parte del processo e il $C_{pm} \rightarrow \infty$ se $\mu \rightarrow T$, come pure se $\sigma \rightarrow 0$, notando quindi come con il C_{pm} i limiti di specifica siano usati solamente per aumentare la deviazione quadratica presente al denominatore.

Quindi in generale può essere osservato come sia il C_p che il C_{pk} diventino arbitrariamente grandi per $\sigma \rightarrow 0$, non dando informazioni su quali valore di T il processo sia centrato, al contrario di C_{pm} che è limitato sia da $d(\mu - T)/3$, e quindi da $\mu \rightarrow T$, quanto da $\sigma \rightarrow 0$.

A questo punto, come per C_p e C_{pk} , viene sottolineata l'importanza che assume **p**, **la percentuale dei prodotti non conformi (NC)**, anche se la motivazione che sta dietro al C_{pm}

non proviene da un esame attento della percentuale NC ma dall'abilità del processo di raggiungere il valore obiettivo. Infatti questo Indice tende ad allontanare l'attenzione dalla adeguatezza del processo alle specifiche, focalizzandosi sulla qualità ottima del prodotto ottenuta solamente quando si raggiunge il valore obiettivo.

Kotz e Johnson (1995) hanno dunque condotto delle ricerche specifiche sulla relazione tra C_{pm} e la percentuale NC, trovando che per un valore dato dell'Indice C_{pm} si otteneva

$$p = \Phi\left(\frac{-d - \mu}{\sqrt{\lambda^2 - \mu^2}}\right) + 1 - \Phi\left(\frac{d - \mu}{\sqrt{\lambda^2 - \mu^2}}\right)$$

dove

$$\lambda = \frac{d}{3 \cdot C_{pm}}$$

e dove

$$d = USL - LSL \text{ dal momento che } USL + LSL = 0$$

3.2.2. Caratteristiche di C_{pm}

Una stima naturale di C_{pm} , \hat{C}_{pm} , è stata trattata da Boyles (1993), focalizzandosi fondamentalmente nel sostituire la quantità

$$E\left(\frac{1}{n} \sum (x_i - T)^2\right) = \sigma^2 + (\mu - T)^2$$

che si trova al denominatore di C_{pm} con un suo stimatore non distorto

$$\tilde{\sigma}(B) = \sqrt{\frac{\sum_{i=1}^n (x_i - T)^2}{n}} = \sqrt{\frac{\sum_{i=1}^n (x_i - T)^2}{n} + (\bar{x} - T)^2} = \sqrt{\frac{n-1}{n} \cdot s^2 + (\bar{x} - T)^2}$$

ottenendo così

$$\hat{C}_{pm(B)} = \frac{USL - LSL}{6\tilde{\sigma}} = \frac{d}{3\tilde{\sigma}}$$

la cui distribuzione è pari a

$$\hat{C}_{pm(B)} \rightarrow \frac{(USL - LSL)\sqrt{n}}{6\sigma\chi_n^2(\lambda)}$$

La costruzione dell'intervallo di confidenza $100(1-\alpha)\%$ per C_{pm} corrisponde alla costruzione dell'intervallo di confidenza per $\sigma^2 + (\mu - T)^2$.

Quindi se usiamo una normale approssimazione per la distribuzione di $\sum(X_i - T)^2$, per grandi valori di n , otteniamo un intervallo $100(1-\alpha)\%$ approssimativamente pari a

$$\frac{1}{n} \sum_{i=1}^n (x_i - T)^2 \pm z_{1-\alpha/2} \cdot \sigma \cdot \left[\frac{2}{\sqrt{n}} \cdot \sqrt{\sigma^2 + 2(\mu - T)^2} \right]$$

dove $\Phi(z_{1-\alpha/2}) = 1-\alpha/2$ per $\sigma^2 + (\mu-T)^2$. Comunque, dal momento che μ e σ non sono conosciuti (altrimenti saremmo a conoscenza del vero valore di C_{pm} e non servirebbe stimarlo), è necessario sostituirli con la loro stima e l'intervallo di confidenza $100(1-\alpha)\%$ che quindi risulterà sarà abbastanza complicato.

A questo punto quindi, si ricavano le stime di

$$\hat{u} = \frac{n \cdot (1 + \hat{\zeta}^2)^2}{1 + 2\hat{\zeta}^2} \quad \text{e} \quad \hat{\zeta} = \frac{\bar{x} - T}{\sqrt{\frac{(n-1) \cdot s^2}{n}}}$$

trovandomi successivamente, per piccoli valori di u , un approssimativo intervallo di confidenza $100(1-\alpha)\%$ per C_{pm} pari a

$$\left(\left(\frac{\chi^2_{1-\alpha, \hat{u}}}{\hat{u}} \right) \cdot \hat{C}_{pm}; \left(\frac{\chi^2_{1-\alpha/2, \hat{u}}}{\hat{u}} \right) \cdot \hat{C}_{pm}; \right)$$

mentre per grandi valori di u ottengo un intervallo di confidenza $100(1-\alpha)\%$ per C_{pm} pari a

$$\left(\left(1 - \frac{z_{1-\alpha/2}}{\sqrt{2\hat{u}}} \right) \cdot \hat{C}_{pm}; \left(1 + \frac{z_{1-\alpha/2}}{\sqrt{2\hat{u}}} \right) \cdot \hat{C}_{pm} \right)$$

3.3. Alcune modifiche dell'Indice C_{pm} .

3.3.1 Gli Indici di Capacità nel caso di tolleranze asimmetriche: C_{pm}^* , C_{pm}^+ e C_{pp}'' .

Come abbiamo visto, l'Indice C_{pm} è stato sviluppato sotto due condizioni di base:

- il processo considerato è distribuito normalmente
- il valore obiettivo si trova alla metà tra i limiti di specifica in $m = \frac{1}{2}(USL + LSL)$

Ci sono molte situazioni in cui queste assunzioni non sono verificate, e questo porta ad una fragilità da parte di C_{pm} , non essendo applicabile in particolare per le situazioni dove il valore obiettivo non è centrato in m .

Per questa ragione è stato studiato da Chan (1988) una generalizzazione di C_{pm} , il C_{pm}^* , così definito per dover trattare anche i processi con tolleranze asimmetriche:

$$C_{pm}^* = \frac{\min(USL - T, T - LSL)}{3\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{d - |T - m|}{3\sqrt{\sigma^2 + (\mu - T)^2}}$$

ottenendo che:

- se $T = m$ allora $C_{pm}^* = C_{pm}$
- se $\mu = m$ allora $C_{pm}^* = C_{pm} = C_{pk}$
- se $T \neq m$ allora il C_{pm}^* è vittima della stessa errata interpretazione che capita a C_{pm} nella stessa situazione

Una stima di C_{pm}^* può essere

$$\hat{C}_{pm}^* = \frac{(d - |T - m|)}{3} \cdot \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - T)^2}$$

notando anche dalla forma come questo Indice sia più chiuso rispetto a C_{pk} e C_{pm} .

Altra modifica rispetto al C_{pm} è l'Indice C_{pm} di **Boyles modificato**, sviluppato come lo dice lo stesso nome da Boyles nel 1995, nel caso in cui si verifichi una situazione di tolleranze asimmetriche in cui $\mu = m \neq T$. Boyles introduce questo Indice da un punto di vista della funzione di perdita quadratica definita come

$$L^*(x) = \left\{ \begin{array}{l} k_1 \cdot (X - T)^2 \leftrightarrow x \leq T \\ k_2 \cdot (X - T)^2 \leftrightarrow x > T \end{array} \right\} \text{ con } k_1 \text{ e } k_2 \text{ entrambe positive}$$

tenendo conto quindi della funzione di perdita asimmetrica. Utilizzando il concetto di perdita relativa attesa che va a quantificare la deviazione dal valore obiettivo e, provvedendo allora alla definizione di un Indice di Capacità focalizzato nel valore obiettivo, Boyles definisce

$$C_{pm}^+ = \frac{1}{3 \sqrt{\frac{1}{(T - LSL)^2} \cdot E_{X < T} [(X - T)^2] + \frac{1}{(USL - T)^2} \cdot E_{X > T} [(X - T)^2]}}$$

dove ho rispettivamente le cosiddette "semi-varianze" (la possibilità che X possa essere uguale a T è ignorata, dal momento che si suppone che si stia trattando una variabile continua, così come $\text{Prob}(X = T) = 0$)

$$E_{X < T} [(X - T)^2] = E[(X - T)^2 | X < T] \cdot \text{Prob}(X < T)$$

e

$$E_{X > T} [(X - T)^2] = E[(X - T)^2 | X > T] \cdot \text{Prob}(X > T)$$

A questo punto si può trovare uno stimatore di questo Indice, stimatore definito come

$$\hat{C}_{pm}^+ = \frac{1}{3 \sqrt{\frac{1}{n} \cdot \left\{ \frac{1}{(T - LSL)^2} \cdot E_{X < T} [(X - T)^2] + \frac{1}{(USL - T)^2} \cdot E_{X > T} [(X - T)^2] \right\}}}$$

Altro Indice costruito per determinarmi la Capacità di processi all'interno dei quali le tolleranze sono asimmetriche è l'Indice C_{pp}'' , derivante da una modifica di C_{pp} . Definiamo per prima cosa il C_{pp} , definito da Greenwich e Jahr-Schaffrath (1995) e considerato come semplice trasformazione di C_{pm}^+ :

$$C_{pp} = \left(\frac{\mu - T}{D} \right)^2 + \left(\frac{\sigma}{D} \right)^2 = C_{ia} + C_{ip}$$

con $D = \min\{(USL - T), (T - LSL)\}/3$, dove C_{ia} è definito come "fattore di inesattezza" e C_{ip} come "fattore di imprecisione".

L'Indice C_{pp} è un Indice che risulta inappropriato se le tolleranze sono asimmetriche, dovendo quindi definire, perché assuma un corretto significato, una sua modifica

$$C_{pp}'' = \left(\frac{A}{D} \right)^2 + \left(\frac{\sigma}{D} \right)^2 = C_{ia}'' + C_{ip}$$

dove $A = \max \left\{ \frac{(\mu - T) \cdot d}{USL - T}, \frac{(T - \mu) \cdot d}{T - LSL} \right\}$.

Per calcolare una stima di C_{pm}'' si tratta il caso in cui la caratteristica del processo sia normalmente distribuita con media μ e varianza σ^2 , ottenendo quindi

$$\hat{C}_{pp}'' = \left(\frac{\hat{A}}{D} \right)^2 + \left(\frac{s}{D} \right)^2$$

con $A = \max \left\{ \frac{(\bar{x} - T) \cdot d}{USL - T}, \frac{(T - \bar{x}) \cdot d}{T - LSL} \right\}$

3.3.2. L'Indice $C_{pm}(a)$.

L'indice C_{pm} tende a rappresentare, come visto precedentemente, una combinazione degli effetti dovuti alla più grande (o più piccola) variabilità e alla più grande (o più piccola) deviazione di μ dal valore obiettivo.

Avendo visto che

$$C_{pm} = \frac{C_p}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

e notando quindi che, se $|\mu - T|/\sigma$ è piccolo, allora

$$C_{pm} \cong \left(1 - \frac{1}{2} \left(\frac{\mu - T}{\sigma}\right)^2\right) \cdot C_p$$

per cercare di bilanciare gli effetti della variabilità e quelli della deviazione dal valore obiettivo, viene introdotto l'Indice $C_{pm}(a)$ così definito

$$C_{pm}(a) \cong \left(1 - a \cdot \left(\frac{\mu - T}{\sigma}\right)^2\right) \cdot C_p$$

dove a è una costante positiva scelta con un criterio particolare per cercare questo bilanciamento. È possibile che $C_{pm}(a)$ assuma anche dei valori negativi oltre che positivi (non come C_p e C_{pm} , ma come C_{pk}).

Uno stimatore naturale di $C_{pm}(a)$ è

$$\hat{C}_{pm}(a) \cong \left(1 - a \cdot \left(\frac{\bar{x} - T}{s}\right)^2\right) \cdot \hat{C}_p = \frac{d}{3s} \left(1 - a \cdot \left(\frac{\bar{x} - T}{s}\right)^2\right)$$

Questo Indice, come C_p , ha il difetto che, se $T \neq \frac{1}{2}(USL + LSL)$, lo stesso valore di $C_{pm}(a)$ ottenuto quando $\mu = T - \delta$ e $\mu = T + \delta$ potrebbe corrispondere ad una differente percentuale NC di pezzi difettosi.

3.3.3. L'indice C_{pmk} .

Come può essere notato, l'Indice C_{pk} è stato ottenuto da una modifica del numeratore di C_p , mentre l'Indice C_{pm} è stato ottenuto modificando il denominatore di C_p .

Pern (1992) ha osservato che combinando il C_{pk} con il C_{pm} si ottiene il C_{pmk} definito come

$$C_{pmk} = \frac{\min(USL - \mu, \mu - LSL)}{3\sqrt{\sigma^2 + (\mu - T)^2}} = \frac{d - |\mu - m|}{3\sqrt{E(X - T)^2}} = \frac{d - |\mu - m|}{3\sqrt{\sigma^2 + (\mu - T)^2}}$$

Come si può notare, al denominatore è stato aggiunto il termine $(\mu - T)^2$ come fattore di penalità dovuta alla deviazione della media del processo dal valore obiettivo.

Questo Indice ha quindi una maggiore sensibilità verso la deviazione dal valore obiettivo rispetto C_{pk} e C_{pm} , ed è capace di distinguere in modo migliore i processi che non raggiungono il valore obiettivo da quelli che lo raggiungono.

Infatti Vamman (1995) ha confrontato il C_{pmk} con il C_p , C_{pm} e con il C_{pk} , trovando, come si può vedere, che questo Indice da maggiore attenzione alla deviazione media del processo dal valore obiettivo rispetto agli altri Indici:

$$C_{pmk} = \left(1 - \frac{|\mu - m|}{d}\right) \cdot C_{pm}$$

$$C_{pmk} = \frac{1}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}} \cdot C_{pk}$$

$$C_{pmk} = \left(1 - \frac{|\mu - m|}{d}\right) \cdot \sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2} \cdot C_p$$

e stipulando infine una classifica in termini di sensibilità alle differenze tra la media del processo e il valore obiettivo così fatta:

1. C_{pmk}
2. C_{pm}
3. C_{pk}
4. C_p

Quindi $C_{pmk} \leq C_{pm}$ e $C_{pmk} \leq C_{pk}$, vedendo che se $\mu = m = T$ allora $C_p = C_{pk} = C_{pm} = C_{pmk}$.

Comunque difficilmente si sa che $\mu = m = T$ e se l'Indice C_{pmk} deve essere stimato perché non sono conosciute la media e la varianza del processo, una sua stima naturale è:

$$\hat{C}_{pmk} = \frac{d - |\bar{x} - m|}{3\sqrt{s^2 + (x_i - T)^2}} = \frac{d - |\bar{x} - m|}{3\sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - T)^2}}$$

Si suppone quindi, additionally alla normalità della distribuzione di X , che $\mu = m = T$, ottenendo come momento r -esimo

$$E(\hat{C}_{pmk}^r) = \frac{1}{3^r \sqrt{n}} \cdot \sum_{j=1}^r (-1)^j \binom{r}{j} \left(\frac{d}{\sigma \sqrt{2}}\right)^{r-j} \cdot \frac{\Gamma\left(\frac{1}{2}(j+1)\right) \Gamma\left(\frac{1}{2}(n-r+j)\right)}{\Gamma\left(\frac{1}{2}(n+j)\right)}$$

È stato trovato da Chen e Hsu (1995), per il caso in cui $\mu = m$, come l'intervallo di confidenza $100(1-\alpha)\%$ per C_{pmk} sia pari a

$$\left(\left(\frac{X_{1-\alpha, \hat{u}}^2}{\hat{u}} \right) \cdot \hat{C}_{pm}; \left(\frac{X_{1-\alpha/2, \hat{u}}^2}{\hat{u}} \right) \cdot \hat{C}_{pm}; \right) \text{ dove } \hat{u} = \frac{n \cdot (1 + \hat{\zeta}^2)^2}{1 + 2\hat{\zeta}^2} \text{ e } \hat{\zeta} = \frac{\bar{x}_u - T}{\hat{\sigma}}$$

Per il caso in cui $\mu = m = T$, come visto, si ottiene $C_{pmk} = C_{pm}$ e quindi per questo caso l'intervallo di confidenza sarà uguale a quello per C_{pm} .

CAPITOLO 4

INDICI DI CAPACITÀ PER DISTRIBUZIONI NON – NORMALI

4.1. Introduzione.

Solo da un breve periodo, sono stati studiati e sono stati messi in atto dei metodi per superare gli effetti della Non-Normalità della distribuzione X sulle proprietà degli Indici di Capacità C_p , C_{pk} , C_{pm} e C_{pmk} .

Kane (1986) in un suo studio, infatti, sostiene: “Una serie numerosa di processi industriali risultano avere una distribuzione Non-Normale per determinate caratteristiche. E’ ragionevolmente probabile aspettarsi che gli Indici di Capacità siano abbastanza sensibili agli scostamenti dalla normalità. Pur tuttavia, è possibile stimare la percentuale di prodotti fuori specifica o direttamente o tramite una distribuzione stimata. Tale percentuale può essere equivalente alla Capacità di un processo avente una distribuzione normale”.

Tutti gli Indici studiati fino ad ora quindi - C_p , C_{pk} , C_{pm} , C_{pmk} - sono stati creati per essere usati con dati distribuiti normalmente. Questa è una assunzione fondamentale per un loro uso corretto e tutte le proprietà statistiche di questi Indici (i loro stimatori e le rispettive proprietà statistiche, gli intervalli di confidenza e la verifica d’ipotesi) dipendono per l’appunto, come sottolinea Gunter (1991), dalla assunzione di normalità “perfetta” dei dati (quindi distribuzioni realmente normali) e non dall’assunzione di normalità “illusiva” (quindi distribuzioni che potrebbero sembrare distribuiti normalmente ma che non lo sono).

Per capire meglio questo, può essere sottolineato come una t di Student simmetrica con molti gradi di libertà (superiore a 30) assomigli quasi perfettamente ad una normale, somiglianza che, se usata, porta però ad una conclusione errata per quanto riguarda il rendimento del processo.

Per fare un esempio che possa sottolineare l'errato uso che provocherebbe l'uso di questi Indici di Capacità nel caso di distribuzioni Non-Normali, supponiamo di considerare un processo perfettamente simmetrico con $C_p = C_{pk} = 1$ ma distribuito secondo una t di Student con sei gradi di libertà. Allora, se considerassimo il processo distribuito normalmente e lo trattassimo da tale, avremmo 2700 parti per milione di pezzi fuori dalle specifiche invece di 10000 parti per milione rispetto a come sarebbe giusto; se invece i dati fossero distribuiti secondo una distribuzione logistica, ce ne sarebbero 82840 di parti per milione fuori dalle specifiche. Quindi, anche se la distribuzione presa è simmetrica come pure indistinguibile da una normale, gli errori del rendimento del processo sarebbero sostanziali.

In questo capitolo quindi si tratterà in prima istanza le proprietà degli Indici di Capacità e le loro stime quando la distribuzione di X ha una forma Non-Normale, per poi successivamente sviluppare alcuni Indici considerati "robusti" alla Non-Normalità.

Questa problematica è direttamente studiata, attraverso diversi metodi, dai seguenti studiosi:

- Kotz e Johnson (1993) hanno sviluppato degli Indici robusti attraverso il metodo di Clememt's, il metodo di Johnson-Pearn-Kotz e i gli Indici di Capacità con distribuzione libera.
- Wright (1995) ha sviluppato l'Indice C_s , derivante da una modifica fatta al denominatore del C_{pmk} in cui si aggiunge un fattore di correzione per l'asimmetria.
- Choi e Bai (1996) ha proposto un metodo basato sulla deviazione standard pesata e sulla varianza standard pesata, cercando così di ottenere un Indice che tiene conto dei gradi di asimmetria tenendo in considerazione separatamente la deviazione standard sopra e sotto la media.
- Vamman ha proposto un Indice generalizzato per dare così una generalizzazione ai quattro Indici C_p , C_{pk} , C_{pm} e C_{pmk} .

Insieme alla determinazione di questi Indici di Capacità cosiddetti "robusti" viene sviluppato anche un metodo per trasformare, attraverso alcune funzioni matematiche, i dati distribuiti Non-Normalmente in dati distribuiti Normalmente. Successivamente a questa trasformazione, viene fatto un test per verificare se la trasformazione fatta è esatta e, se è così, allora posso usare gli Indici di Capacità sviluppati nel caso di distribuzioni Normali.

Per quanto sottolineare come nella realtà non sia difficile trovare queste situazioni di Non-Normalità, Gunter (1989) in un suo articolo ("The Use and Abuse of C_{pk} ") descrive tre situazioni particolari di non normalità e a queste situazioni associa dei processi industriali reali. Queste tre situazioni sono: 1) le distribuzioni asimmetriche limitate solo da un lato; 2) le distribuzioni che hanno delle code pesanti dove la deviazione dal controllo statistico causa una minor consistenza rispetto alla Normale; 3) le distribuzioni con le code corte dove i dati sono distribuiti più o meno tra i limiti di specifica.

La prima situazione accade per esempio con le misurazione delle dimensioni di superfici ruvide o ovalizzate, di strati spessi e di tutto quello che non può essere più piccolo di zero. In questo caso i dati del processo si raggruppano quindi intorno al valore più basso, in questo casi lo zero, mentre la coda si stende lungamente a destra.

La seconda situazione sottolinea come le distribuzioni Normali siano centrate intorno al loro punto centrale ma come, per molti processi industriali, questo sia praticamente impossibile, essendo infatti inevitabili le cause cosiddette 'accidentali' e non 'identificabili' che portano il processo al limite o lontano dalla sua centralità. Risultato che può anche essere ottenuto se il processo tende a operare la maggior parte del suo tempo in controllo statistico e occasionalmente, ma regolarmente, qualcosa diventa meno consistente. È molto difficile poter scoprire questi effetti, dal momento che si richiede una gran quantità di dati derivanti da un processo stabile.

La terza situazione è quella della verifica di dove il processo è troppo variabile, dovendo quindi esaminare quei prodotti difettosi che stanno nelle code. Un esempio tipico di processi industriali che rispecchiano questa situazione si ha quando qualcuno assembla gli output da macchine o linee di produzione differenti.

Quindi, dalla descrizione di queste tre situazioni, può essere concluso come le distribuzioni Non-Normali siano qualcosa di frequente nella realtà industriale e come ci sia la necessità di trovare degli Indici di Capacità appropriati, dal momento che gli Indici C_p , C_{pk} , C_{pm} , C_{pmk} non possono essere usati correttamente.

4.2. Gli Effetti della Non – Normalità.

4.2.1. Introduzione.

Per osservare gli effetti della Non-Normalità, Gunter (1989) ha studiato l'interpretazione del C_{pk} sotto tre differenti distribuzioni Non-Normali. Tali distribuzioni hanno tutte la medesima media μ e deviazione standard σ , avendo pertanto i medesimi valori di C_p e C_{pk} .

Le tre distribuzioni sono:

1. Una distribuzione asimmetrica con un limite inferiore finito chiamata $\chi_{4,5}^2$ (Chi – Quadro con 4.5 gradi di libertà)
2. Una distribuzione leptocurtica ($\beta_2 > 3$) chiamata distribuzione t_s
3. Una distribuzione uniforme

Riprendendo le proporzioni in ppm di valori non conformi e ricordando che una distribuzione normale prevede in media 2700 ppm metà sopra il limite $+3\sigma$ e metà sotto il limite -3σ , le proporzioni in ppm oltre i limiti di $\pm 3\sigma$ per le tre distribuzioni sono approssimativamente:

- per la distribuzione (1) 14000 ppm (tutti superiori a $\pm 3\sigma$)
- per la distribuzione (2) 4000 ppm (metà per valori superiori a $+3\sigma$ e metà per valori inferiori a -3σ)
- per la distribuzione (3) zero ppm

Nella figura1 sono rappresentate le tre distribuzioni appena citate con l'aggiunta di una normale, sono tutte standardizzate con media $\mu = 0$ e deviazione standard $\sigma = 1$.

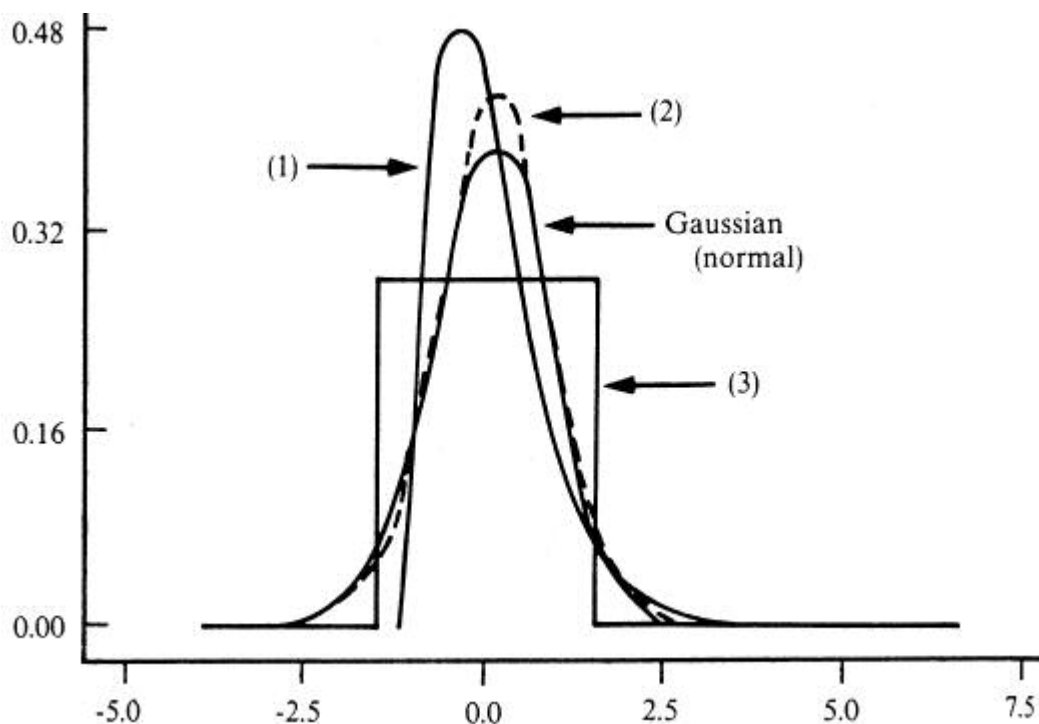


Figura 1

Altri studiosi che hanno studiato gli effetti della Non-Normalità sugli stimatori degli Indici di Capacità di Base, sono stati English e Taylor (1990), i quali hanno fatto una simulazione di Monte Carlo della distribuzione della stima di C_p (con $C_p = 1$, $\mu = 0$, $\sigma = 1$, $LSL = -3$ e $USL = 3$) per una distribuzione Normale, una triangolare simmetrica, una uniforme ed esponenziale, considerando la numerosità campionaria n pari a 5, 10, 20, 50. Da questa simulazione e dai valori di $\text{Prob}(\hat{C}_p \geq c)$ ottenuti, hanno osservato per tutte le distribuzioni Normali e Non-Normali, la rilevante differenza tra il vero valore di C_p e la sua stima quando n è minore di 20. Più importante ancora è osservare le grandi differenze che ci sono, specialmente per grandi valori di c , tra le distribuzioni asimmetriche, come l'esponenziale, e quelle simmetriche.

Kotz e Johnson (1993) infatti hanno stabilito che il problema fondamentale della non-normalità era quello di portare i ricercatori ad esaminare le proprietà di particolari Indici di Capacità attraverso la valutazione dei loro stimatori per una grande varietà di processi distribuiti Non-Normalmente. Dal momento che il numero di distribuzioni Non-Normali è grande, un approccio ovvio è quello di scegliere il momento terzo standardizzato (**l'asimmetria**) e il momento quarto standardizzato (**la curtosi**). Il primo cattura la deviazione dalla simmetria mentre il secondo cattura il peso delle code. Infatti una asimmetria positiva indica una coda lunga a destra mentre una asimmetria negativa indica una coda lunga a sinistra. Per quanto riguarda la curtosi invece, una curtosi positiva indica una curva appuntita mentre una curtosi negativa indica una curva più piatta rispetto la normale.

Di sicuro, l'uso dei momenti è una limitazione, dal momento che la specificazione dei primi quattro momenti non è sufficiente per determinare una distribuzione, ma comunque un loro studio porta a dei significativi risultati per la risoluzione del problema.

Un particolare vantaggio nell'usare i momenti, è quello che la gran parte delle distribuzioni può essere campionata attraverso l'uso di famiglie di distribuzioni, come quella di Pearson, di Burr, di Johnson e la distribuzione di Edgeworth, tutte famiglie che possono essere determinate attraverso il momento primo.

Se si vuole quindi un calcolo preciso degli Indici di Capacità in presenza di distribuzioni campionarie Non-Normali, occorre possedere delle stime della media, della deviazione standard, dell'asimmetria e della curtosi del processo. Questi valori statistici potranno essere calcolati dall'insieme dei dati totali per un processo stabile, ottenendo per il parametro dell'asimmetria, $\sqrt{\beta_1}$, e per il parametro della curtosi, β_2 , le rispettive relazioni:

$$\sqrt{\beta_1} = \frac{n \cdot \sum_{i=1}^n (x_i - \bar{x})^3}{(n-1) \cdot (n-2) \cdot s^3}$$

$$\beta_2 = \frac{n \cdot (n+1) \sum_{i=1}^n (x_i - \bar{x})^4}{(n-1) \cdot (n-2) \cdot (n-3) \cdot s^4} - \frac{3 \cdot (n-1)^2}{(n-2) \cdot (n-3)}$$

Tali parametri diventano poco significativi se la numerosità campionaria è molto elevata, potendo quindi calcolarli utilizzando la maggior parte di software statistici (come per esempio Statgraphics)

Osservando maggiormente l'aspetto teorico, Kocherlakota (1992) ha quindi stabilito la distribuzione di $\hat{C}_p = d/3$ per due casi particolari:

- a) quando la distribuzione del processo è una normale contaminata con $\sigma_1 = \sigma_2 = \sigma$ (una incrocio tra due distribuzioni normali con differenti medie ma le stesse varianze)
- b) quando la distribuzione del processo è una distribuzione di Edgeworth (una serie che coinvolge la normale standard e le sue derivate) che ha:

$$f_x(x) = \left(1 - \frac{1}{6} \lambda_3 D^3 + \frac{1}{24} \lambda_4 D^4 + \frac{1}{72} \lambda_3^2 D^6 \right) \cdot \varphi(x; 0; 1)$$

dove D^j è la derivata j-esima rispetto a X e $\lambda_3 = \sqrt{\beta_1}$ e $\lambda_4 = \beta_2 - 3$ sono rispettivamente la misura standardizzata dell'asimmetria e della curtosi.

4.2.2. Il caso della Normale Contaminata.

Kotz e Johnson (1993) sono riusciti a calcolare i momenti di \hat{C}_p per una normale contaminata con k componenti (componenti che hanno tutte la stessa varianza σ^2) e con funzione di ripartizione pari a

$$\sum_{j=1}^k p_j \cdot \varphi(x; \mu_j; \sigma)$$

Un campione casuale con numerosità campionaria n e tratto da una popolazione con la precedente funzione di ripartizione, può essere visto come un miscuglio tra campioni causali di numerosità N_1, \dots, N_k tratti da popolazioni con funzioni di ripartizione $\varphi(x, \mu_1, \sigma), \dots, \varphi(x, \mu_k, \sigma)$ (dove $\varphi(x, \mu_i, \sigma)$ è la densità di una normale con parametri μ_i e σ). Può essere dimostrato come la numerosità $N = (N_1, \dots, N_k)$ abbia una distribuzione multidimensionale con parametri $(n; p_1, \dots, p_k)$, ottenendo così che

$$\text{Pr ob}(N = n) = \text{Pr ob} \left(\bigcap_{j=1}^k (N_j = n_j) \right) = \frac{n!}{\prod_{j=1}^k n_j!} \prod_{j=1}^k p_j^{n_j}$$

con $\sum_{j=1}^k n_j = n$ e con $\sum_{j=1}^k p_j = 1$

A questo punto si scopre che la distribuzione condizionale della statistica è

$$s'^2 = \sum_{j=1}^k (x_j - \mu_0)^2 \approx \sigma^2 \times \chi_{n-1}^2(\lambda_n^*)$$

con parametro di non centralità pari a

$$\lambda_n^* = \sum_{j=1}^k n_j \cdot (\mu_j - \mu_0)^2$$

Quindi, per il campione preso in considerazione ottengo che

$$s^2 = \sum_{j=1}^k (x_j - \bar{x})^2 \approx \sigma^2 \times \chi_{n-1}^2(\lambda_n)$$

con parametro di non centralità pari a

$$\lambda_n = \sum_{j=1}^k n_j \cdot (\mu_j - \bar{\mu}_n)^2$$

Condizionata a $\mathbf{N} = \mathbf{n}$, \hat{C}_p si distribuisce come

$$\hat{C}_p = \frac{d}{3\sqrt{\frac{s^2}{n-1}}} \approx \frac{d\sqrt{n-1}}{3\sigma \cdot \chi_{n-1}^2(\lambda_n)}$$

e, riprendendo la formula dei momenti della distribuzione Chi-Quadro non centrata, si ottiene che il momento r-esimo di \hat{C}_p è:

$$E[\hat{C}_p^r] = \sum_n \frac{n!}{k \prod_{j=1}^k n_j!} \prod_{j=1}^k p_j^{n_j} \cdot E(\hat{C}_p^r | n)$$

Kotz e Johnson (1993) a questo punto hanno calcolato una tabella all'interno della quale sono espressi i valori di $E(\hat{C}_p) | C_p$ e di $DS(\hat{C}_p) | C_p$ in funzione di \mathbf{a} e \mathbf{p} , dove $\mu_1/\sigma = -a$, $\mu_2=0$, $\mu_3/\sigma = a$ e $p_1 = p_3 = p$.

Da questa tabella si nota come:

- a) Per un dato valore di n , la distorsione e la deviazione standard decrescono entrambe quando n aumenta
- b) Per un dato valore di a , la distorsione e la deviazione standard decrescono entrambe quando n aumenta
- c) Per un dato valore di a e n , la distorsione di a e la deviazione standard non variano con p_1 e p_3 quando $k = 3$ o con p quando $k = 2$. La distorsione è numericamente più grande quando i parametri cosiddetti "contaminati" sono piccoli
- d) D'altro canto, più grande è il valore di a , più marcata è la varianza rispetto a p_1 e a p_3 o p .
- e) La distorsione è positiva quando a è piccolo ma decresce con l'aumentare della percentuale di contaminazione.

Comunque la distorsione è negativa per grandi valori di a e p e diventa molto spiccata quando n aumenta. Questo è il contrario di cosa accade quando $p = 0$ e quindi quando non c'è nessuna contaminazione, caso in cui la distorsione della stima di C_p è sempre positiva, sebbene diminuisca quando n aumenta.

4.2.3. Il caso della distribuzione di Edgeworth.

L'uso della distribuzione di Edgeworth per rappresentare lievi deviazioni dalla normalità è stato in questi ultimi anni un metodo che è diventato sempre più frequente, e il pioniere nel suo uso è stato Subrahmaniam (1966).

Deve essere tenuto in mente che, per assicurare che la funzione di ripartizione rimanga positiva per tutti i valori di x , ci sono delle limitazioni abbastanza severe sui valori ammissibili di $\sqrt{\beta_1}$ e β_2 .

Comunque Kocherlakota (1992) ha dimostrato che il valore atteso di \hat{C}_p per una distribuzione di Edgeworth è

$$E[\hat{C}_p] = \frac{\sqrt{n-1} \cdot \Gamma\left(\frac{1}{2}(n-2)\right)}{\sqrt{2} \cdot \Gamma\left(\frac{1}{2}(n-1)\right)} \cdot C_p \left\{ 1 + \frac{3}{8} g_n (\beta_2 - 3) - \frac{3}{8} h_n \beta_1 \right\}$$

dove

$$g_n = \frac{n-1}{n(n+1)} \text{ e } h_n = \frac{(n-2)}{n(n+1)(n+3)}$$

Kocherlakota (1992) allo stesso tempo, dopo aver determinato il valore atteso della stima di C_p , ha tabulato la varianza della stima di C_p , varianza che risulta pari a

$$\text{Var}[\hat{C}_p] = \frac{n-1}{n(n-3)} \cdot C_p^2 \left[1 + \frac{1}{n} \left\{ 1 + g_n (\beta_2 - 3) - \frac{1}{9} h_n \beta_1 \right\} \right]$$

In questa tabulazione si sono ricavati i valori di $E(\hat{C}_p)|C_p$ e di $DS(\hat{C}_p)|C_p$ per determinati valori di $\sqrt{\beta_1}$ e β_2 . Si è notato come per un dato valore di β_2 il valore della media non vari con $\sqrt{\beta_1}$ e come invece il valore della deviazione standard vari solamente poco con $\sqrt{\beta_1}$. Lo stesso vale per un dato valore di $\sqrt{\beta_1}$, infatti il valore della deviazione standard aumenta con β_2 . Tutto questo sta quindi a sottolineare la robustezza della stima di C_p verso una asimmetria lieve.

Kocherlakota (1992), sempre trattando questa parte, ha anche studiato la distribuzione di uno stimatore naturale di $C_{pu} = (USL - \mu)/3\sigma$ trovando

$$\hat{C}_{pu} = \frac{USL - \bar{x}}{3s}$$

A questo punto è stata fatta una tabella all'interno della quale viene calcolata, come nei casi precedenti, la media e la deviazione standard della stima di C_{pu} , osservando quindi che, se il processo risulta simmetrico e quindi $\sqrt{\beta_1} = 0$, $E(\hat{C}_p)|C_p$ è proporzionale a C_{pu} . Si osserva pure che, quando $\sqrt{\beta_1} \neq 0$, la media della stima di C_{pu} è veramente quasi proporzionale a C_{pu} , tenendo però presente che, quando β_1 è grande, non è possibile rappresentare la distribuzione nella forma di Edgeworth.

4.2.4. Altri casi.

Price e Price nel 1992 hanno presentato i valori, provenienti da una simulazione, del valore atteso di \hat{C}_p e \hat{C}_{pk} per tredici distribuzioni. Queste distribuzioni sono una Normale, una Uniforme, una esponenziale, due Beta e otto Gamma, ordinate in ordine crescente secondo il loro coefficiente di asimmetria $\sqrt{\beta_1}$ ma tutte con lo stesso valore atteso pari a 50 e deviazione standard pari a 1.

Distribuzione	Asimmetria $\sqrt{\beta_1}$
Normale (50, 1)	0
Uniforme (48.268, 51.732)	0
10 x Beta (4.4375, 13.3125)+47.5	0.506
10 x Beta (13.3125, 4.4375)+42.5	-0.506
Gamma (9, 3)+47	0.667
Gamma (4, 2)+48	1
Gamma(2.25, 1.5)+48.5	1.333
(Esponenziale) Gamma(1, 1) +49	2
Gamma(0.75, 0.867)+49.1340	2.309
Gamma(0.5, 0.707)+49.2929	2.828
Gamma(0.3, 0.5477)+49.3675	3.163
Gamma(0.3, 0.5477)+49.4523	3.651
Gamma(0.25, 0.5)+49.5	4

Si è successivamente osservato come, per C_p , il valore di $E(\hat{C}_p)|C_p$ non dipenda da μ e non dipenda nemmeno da C_p . È stata successivamente determinata una tabella in cui si calcola questo valore atteso per le tredici distribuzioni, sempre ordinate secondo l'ordine di $|\sqrt{\beta_1}|$. Analizzando inizialmente queste stime per la Normale, si osserva come esse siano superiori rispetto al valore esatto del 2% per $n = 10$, del 0.5% per $n = 30$ e del 0.2% per $n = 100$. Analizzando in seguito le distribuzioni delle Gamma, soprattutto per $n = 100$, è evidente come anche per queste ci siano delle variazioni campionarie in ordine progressivo. Infatti, quando l'asimmetria aumenta, il valore atteso di \hat{C}_p aumenta, raggiungendo valori

abbastanza ragguardevoli per alti valori di $\sqrt{\beta_1}$. Questo comunque corrisponde a distribuzioni che hanno una forma abbastanza pesante, un indice di asimmetria $\sqrt{\beta_1}$ più grande di quello dell'esponenziale. Comunque per moderati valori di $\sqrt{\beta_1}$, il valore atteso della stima di C_p è abbastanza racchiuso attorno a quello della distribuzione normale.

A questo punto può anche essere trattato il valore atteso di \hat{C}_{pk} derivante sempre dalle tredici distribuzioni determinate da Price e Price. È stato osservato come ci sia, per l'ultima distribuzione Gamma, una distorsione positiva grande per $n = 10$ mentre, per $n = 30$ e per $n = 100$, la distorsione diventi piccola. Anche per la distribuzione Esponenziale c'è sempre una distorsione positiva in \hat{C}_{pk} , distorsione che diventa più grande quando μ supera il valore di $\frac{1}{2}(USL + LSL)$ rispetto a quando ne è inferiore. Questa distorsione va dal 25% - 30% quando $n = 10$ al 2.5% - 5% quando $n = 100$. Comunque, come per \hat{C}_p , i risultati ottenuti per l'ultima distribuzione Gamma sono molto differenti rispetto a quelli ottenuti dalla Normale e dalle distribuzioni con asimmetria più piccola.

Spostandoci ora alla variabilità di \hat{C}_{pk} e di \hat{C}_p , si nota come la deviazione standard di C_p possa essere considerata approssimativamente proporzionale a $\sqrt{(\beta_2 - 1)}$, dove $\beta_2 = \mu_4/\sigma^4$ e dove rappresenta il fattore della curtosi per la distribuzione del processo. Ci aspetteremo quindi delle deviazioni più basse per i processi con distribuzione uniforme ($\beta_2 = 1.8$) rispetto a quelli con una distribuzione Normale ($\beta_2 = 3$), mentre ci aspetteremo delle deviazioni più grandi per le distribuzioni con parametro $\beta_2 > 3$, come per esempio l'esponenziale.

4.3. Trasformazione dei dati distribuiti Non-Normalmente in dati distribuiti normalmente: "il metodo di Johnson".

4.3.1. Introduzione.

Come visto fin dall'inizio, molte delle tecniche usate per determinare se il processo è in controllo statistico, derivano dallo studio di caratteristiche che si distribuiscono in modo Normale. Allo stesso tempo si è osservato come esistano molti processi industriali distribuiti in modo Non-Normale, portando quindi questa deviazione dalla normalità seri problemi nelle proprietà degli stimatori. Il modo più semplice per trattare quindi i dati distribuiti in modo Non-Normale è quello di trasformare i dati, attraverso alcune funzioni matematiche, in dati distribuiti normalmente, applicando successivamente le procedure di controllo della qualità sui dati trasformati. Un metodo molto usato è quello sviluppato da Johnson, che si basa nel trasformare i dati attraverso **il metodo delle tre trasformazioni di Johnson S_L , S_U , S_B** .

Come sottolineato da Somerville e Montgomery (1996), ci sono vari approcci principali per trattare i dati distribuiti in modo Non-Normale, e il più importante è quello che cerca di trasformare questi dati attraverso una generica famiglia di distribuzioni, famiglia spesso con quattro parametri da stimare. Le tre famiglie di distribuzioni più usate sono la distribuzione λ generalizzata di Ramberg (1979), la distribuzione di Pearson per la quale sono state create delle tabelle da parte di Johnson, Nixon e Amos (1963) e per finire la più importante distribuzione di Johnson (1949). Quindi, l'approccio più importante è quello che si basa sostanzialmente sulla famiglia delle distribuzioni di Johnson, ed ha due vantaggi fondamentali: il primo è quello di essere semplice da usare e secondo, chi analizza il processo considerato, può intuire quali sono le sue caratteristiche cercando di adattare la distribuzione attraverso delle funzioni particolari.

In prima istanza, Shapiro e Haman (1967) scoprirono che la scelta su quale funzione usare si basasse sul fatto che, in un plot del momento terzo e quarto standardizzato ($\sqrt{\beta_1}; \beta_2$), la distribuzione S_L formasse una curva che divideva il piano ($\sqrt{\beta_1}, \beta_2$) in due regioni. Quindi, trovate queste due regioni, si scoprì che la distribuzione S_U cadeva in una delle due regioni mentre la distribuzione S_B cadeva nell'altra regione.

Successivamente a questo, nell'uso del sistema di Johnson, il primo passo da fare era quello di determinare quale delle tre famiglie era più consona ai dati del processo preso in considerazione. La procedura più comune era quella di calcolare le stime campionarie dei momenti terzo e quarto standardizzati, scegliendo successivamente la distribuzione associata a quella regione in cui i dati calcolati cadevano dentro.

I maggiori difetti di questa procedura però sono:

1. le varianze delle stime del momento terzo e quarto sono molto alte.
2. le stime dei tre momenti sono molto distorte per piccoli campioni.

3. gli stimatori dei momenti sono molto condizionati dagli altri.

Una volta che la famiglia è stata scelta, il passo successivo è quello di determinare i parametri. Le procedure per la stima dei parametri di S_L sono state fatte inizialmente da Aitchison e Brown (1957). Per quanto riguarda S_B invece, Johnson (1947) suggerì delle tecniche che si basavano, nel caso in cui uno o entrambi i limiti fossero conosciuti, nel cercare i percentili della distribuzione di S_B dando delle formule matematiche, mentre nel caso in cui nessuno dei due limiti fosse conosciuto, in procedure che richiedevano la soluzione di equazioni non lineari. Allo stesso tempo, per quanto riguarda S_U , Johnson suggerì una tecnica basata nel confrontare i momenti della distribuzione di S_U , tecnica che richiedeva una tabella a doppia entrata per determinare i parametri γ e η derivanti dalla stima campionaria di $\sqrt{\beta_1}$ e β_2 . Infine Burak (1972) derivò una equazione di quarto grado per i parametri di S_B basata sulle quattro deviazioni normali simmetriche ed equidistanti, mentre Mage (1980) ridusse queste equazioni di quarto grado in una equazione quadratica che portava a esplicite soluzioni dei quattro parametri da determinare.

Quindi, dopo questa prima fase, Shapiro e Slikfer (1980) decisero di introdurre un metodo semplice di selezione, metodo che usa una funzione dei quattro percentili per selezionare quale delle tre famiglie è più adatta al processo preso in considerazione, dando quindi una stima dei parametri per tutte le famiglie senza richiedere una soluzione di equazioni simultanee o di speciali tabelle.

In seguito allo sviluppo di Shapiro e Slikfer, Chou, Polansky e Mason (1998), hanno presentato un algoritmo che, mentre per la stima dei parametri usa sempre la teoria dei percentili, per la stima su quale delle tre trasformazioni di Johnson sia corretta per i dati presi in considerazione, usa principalmente il criterio del test di Shapiro-Wilk

È stato scoperto come questo algoritmo sia abbastanza usato nelle applicazioni del Controllo Statistico della Qualità e come siano stati rari i casi in cui, sebbene esistano, alcune serie di dati non possano essere trasformate attraverso questa procedura in serie distribuite normalmente. È bene sottolineare come questo algoritmo sia adatto per cercare la migliore trasformazione di Johnson, nel senso che la trasformazione risultante dei dati è la più vicina alla Normalità. Infine è importante ricordare quindi che le trasformazioni che questo algoritmo produce sono specifiche per i valori osservati, infatti se nuovi dati venissero introdotti allora bisognerebbe fare una nuova trasformazione.

4.3.2. Le trasformazioni di Johnson.

Johnson ha sviluppato un sistema di distribuzioni basato sul metodo dei momenti, proprio come Pearson. La formula generale della trasformazione di Johnson è data da:

$$Z = \gamma + \eta k(x; \epsilon, \lambda)$$

con $\eta > 0$; $\lambda > 0$; $-\infty < \gamma < \infty$; $-\infty < \epsilon < \infty$, dove Z è una variabile casuale Normale e X è la variabile casuale che deve essere trasformata dalla distribuzione di Johnson.

I quattro parametri γ , η , ϵ e λ sono stimati mentre k è una funzione arbitraria che può assumere una delle seguenti distribuzioni:

- a) Una distribuzione lognormale (S_L) pari a $k_1(x; \epsilon; \lambda) = \log\left(\frac{x - \epsilon}{\lambda}\right)$ con $x \geq \epsilon$.
- b) Una distribuzione non limitata (S_U) pari a $k_2(x; \epsilon; \lambda) = \sinh^{-1}\left(\frac{x - \epsilon}{\lambda}\right)$ con $-\infty < x < \infty$.
- c) Una distribuzione limitata (S_B) pari a $k_3(x; \epsilon; \lambda) = \log\left(\frac{x - \epsilon}{\lambda + \epsilon - x}\right)$ con $\epsilon \leq x \leq \epsilon + \lambda$.

Il sistema proposto da Johnson quindi contiene tre famiglie di distribuzioni che sono generate dalle trasformazioni seguenti

- a) la trasformazione lognormale S_L pari a $Z = \gamma + \eta \cdot \log\left(\frac{x - \epsilon}{\lambda}\right)$.
- b) la trasformazione non limitata S_U pari a $Z = \gamma + \eta \cdot \sinh^{-1}\left(\frac{x - \epsilon}{\lambda}\right)$.
- c) la trasformazione limitata S_B pari a $Z = \gamma + \eta \cdot \log\left(\frac{x - \epsilon}{\lambda + \epsilon - x}\right)$.

Per quanto riguarda la lognormale S_L , questa è una lognormale a tre parametri e non a quattro come potrebbe sembrare, dal momento che il parametro λ può essere eliminato sostituendo il parametro γ con $\gamma^* = \gamma - \eta \cdot \log(x - \epsilon)$.

4.3.3. Procedure di selezione e stima dei parametri.

Questo studio sulle procedure di selezione e di stima dei parametri, è motivato dal tentativo di trovare una proprietà della trasformazione $Z = \gamma + \eta k(x; \epsilon; \lambda)$ che possa essere usata per selezionare l'appropriata distribuzione della famiglia di Johnson, approssimando quindi una serie di dati che non richiedono l'uso di momenti di ordine elevato.

Iniziamo quindi considerando una qualsiasi distribuzione descritta dall'equazione $Z = \gamma + \eta k(x; \epsilon; \lambda)$. Scelto quindi un valore fissato $z > 0$ di una Normale standard, si determinano successivamente i quattro punti $-3z, -z, z, 3z$ trovando così tre intervalli di uguale lunghezza, osservando come la trasformazione Z presa in considerazione produca quattro valori di x che non sono distanziati equamente (generalmente comunque, i quattro punti $(-3z, -z, z, 3z)$, che sono simmetrici intorno al punto 0, sono un'appropriata scelta se l'intera serie dei dati è stata adattata da una distribuzione di Johnson). Si ipotizza quindi che, per una distribuzione di Johnson limitata, la distanza tra ognuno dei punti esterni ed interni sia più piccola della

distanza tra i due punti interni, e che sia vero il contrario per una distribuzione non limitata. Questo è facile da dimostrare e porta al seguente risultato più generale.

Determinato che, per una normale standard, (p_1, p_2, p_3, p_4) sono le aree sotto i punti $(-3z, -z, z, 3z)$, si ottengono sotto la trasformazione Z i valori dei corrispondenti quantili $x_{3z}, x_z, x_{-z}, x_{-3z}$, calcolando quindi:

$$m = x_{3z} - x_z$$

$$n = x_{-z} - x_{-3z}$$

$$p = x_z - x_{-z}$$

A questo punto si definisce il “rapporto quantile RQ ” definito come

$$RQ = \frac{m \cdot n}{p^2}$$

e si ricava, basandosi sul rapporto quantile RQ , il seguente criterio per determinare quale delle tre trasformazioni di Johnson è più appropriata per il processo preso in considerazione:

$RQ > 1$ per ogni distribuzione **S_U**

$RQ < 1$ per ogni distribuzione **S_B**

$RQ = 1$ per ogni distribuzione **S_L**

Riprendendo le tre trasformazioni di Johnson e ricordando che la distribuzione **S_U** è non limitata, la distribuzione **S_L** è limitata da un lato e la distribuzione **S_B** è limitata da entrambi i lati, si nota come i pedici U, L e B si riferiscano alla natura del campo di variabilità della distribuzione. Bisogna quindi fare molta attenzione, come sottolinea Chou (1994), ad una cosa fondamentale: “data una distribuzione limitata da un lato, questa non appartiene necessariamente alla famiglia lognormale, tuttavia solamente se lo fosse allora $RQ = 1$ ”.

4.3.3.1. Procedura di selezione.

La procedura di selezione consiste nei seguenti passi:

1. Si sceglie un valore di $z > 0$. Questa scelta potrebbe essere motivata dal numero dei dati. In generale per un numero moderato di dati si sceglie un valore di z inferiore a uno. Un valore di z pari a uno o più grande renderebbe difficile la stima del percentile corrispondente a $\pm 3z$. Una scelta più “normale” è quella di usare un valore di z vicino a 0.5, per esempio 0.526, imponendo quindi $\pm 3z = \pm 1.572$ e facendo così in modo che questi due punti portino a stimare il 70-esimo e 94.5-esimo percentile. Comunque, più elevato è il numero di osservazioni e più elevato è il valore di z che deve essere scelto, potendo quindi concludere che, per la scelta di z , si possa mirare verso la scelta di un valore che assuma una distribuzione empirica chiusa nell’area di

più grande interesse. Può anche infine essere dimostrato come z e le stime di p , m , n possano essere usate per ottenere le stime dei parametri delle distribuzioni.

2. Si determina, dalla tavola dei valori della funzione di ripartizione $\Phi(z)$ della Normale standardizzata, la percentuale P_ζ con ζ che corrisponde rispettivamente a $(-3z, -z, z, 3z)$. Per esempio se $z = 0.2$ allora $P_{0.2} = 0.5793$.
3. Per ogni ζ , si ottiene dai dati il percentile $x^{(i)}$ corrispondente a $P_\zeta = (i - 1/2)/n$ con $i = 1, 2, 3, 4$ e con n uguale al numero di dati, potendo concludere di conseguenza che $x^{(i)} = x_\zeta$. A questo punto si osserva come x_ζ corrisponda all' i -esima osservazione ordinata, dove $i = nP_\zeta + 1/2$. Dal momento che i , in generale, non è un integrale, sarà necessariamente interpolato.
4. Dai valori ottenuti nel passo precedente si calcolano i valori di m , n , p e si usa il criterio RQ per selezionare la trasformazione più appropriata.

Si può notare infine come, dal momento che $x^{(i)}$ sono variabili casuali, la probabilità che $RQ=1$ è nulla. Se si spera di usare la trasformazione S_L , ci si deve rendere conto che è necessario ottenere un intervallo di tolleranza intorno a uno.

4.3.3.2 Stima dei parametri.

Dopo che il processo di selezione è stato completato, il successivo problema è quello di stimare i parametri della distribuzione scelta. Come spiegato nella parte introduttiva, varie tecniche di stima dei parametri sono state, ma l'approccio più comune è quello basato sui percentili. Grazie a questa, sono state infatti date delle formule per ognuna delle tre distribuzioni, formule che possono essere usate facilmente per determinare la stima dei parametri senza l'aiuto delle tavole. Queste stime sono fatte in termini di un valore prescelto di z e dei valori assunti da m , n , p . Infatti, per ognuna delle tre famiglie, le formule sono ottenute partendo da una data distribuzione di Johnson e da un valore fissato di $z > 0$, e successivamente risolte in termini di z , m , n , p .

a) Distribuzione non limitata di Johnson S_U

La Distribuzione non limitata di Johnson S_U è

$$Z = \gamma + \eta \cdot \log\left(\frac{x - \epsilon}{\lambda}\right)$$

e i valori dei parametri per questa distribuzione sono presentati cercando di enfatizzare la loro dipendenza da m/p e n/p . La stime dei parametri sono quindi:

$$\eta = \frac{2z}{\operatorname{conh}^{-1}\left[\frac{1}{2}\left(\frac{m}{p} + \frac{n}{p}\right)\right]} \quad \text{con } \eta > 0$$

$$\gamma = \eta \sinh^{-1} \left[\frac{\frac{n-m}{p-p}}{2 \sqrt{\frac{m}{p} \cdot \frac{n}{p} - 1}} \right]$$

$$\lambda = \frac{2p \cdot \sqrt{\frac{m}{p} \cdot \frac{n}{p} - 1}}{\left(\frac{m}{p} + \frac{n}{p} - 2\right) \sqrt{\frac{m}{p} + \frac{n}{p} - 2}} \quad \text{con } \lambda > 0$$

$$\varepsilon = \frac{x_z - x_{-z}}{2} + \frac{p \left(\frac{n-m}{p-p}\right)}{2 \left(\frac{m}{p} + \frac{n}{p} - 2\right)}$$

b) Distribuzione limitata di Johnson S_B

La Distribuzione limitata di Johnson S_B è

$$Z = \gamma + \eta \cdot \log \left(\frac{x - \varepsilon}{\lambda + \varepsilon - x} \right)$$

e la soluzione per i parametri di questa distribuzione sottolinea la loro dipendenza dal rapporto p/m e p/n . La stime dei parametri sono quindi:

$$\eta = \frac{z}{\cosh^{-1} \left[\frac{1}{2} \sqrt{\left(1 + \frac{p}{m}\right) \left(1 + \frac{p}{n}\right)} \right]} \quad \text{con } \eta > 0$$

$$\gamma = \eta \sinh^{-1} \left[\frac{\left(\frac{p-p}{n-m}\right) \sqrt{\left(1 + \frac{p}{m}\right) \left(1 + \frac{p}{n}\right) - 4}}{2 \left(\frac{p}{m} \frac{p}{n} - 1\right)} \right]$$

$$\lambda = \frac{p \sqrt{\left(\left(1 + \frac{p}{m}\right) \cdot \left(1 + \frac{p}{n}\right) - 2\right)^2 - 4}}{\frac{p}{n} \frac{p}{m} - 1} \quad \text{con } \lambda > 0$$

$$\varepsilon = \frac{x_z - x_{-z}}{2} - \frac{\lambda}{2} + \frac{p \left(\frac{p-p}{n-m}\right)}{2 \left(\frac{p}{m} \frac{p}{n} - 1\right)}$$

c) Distribuzione non limitata di Johnson S_L

La Distribuzione lognormale di Johnson S_L è

$$Z = \gamma^* + \eta \cdot \sinh^{-1}(x - \varepsilon)$$

e per stimare questi parametri fondamentalmente si usa il metodo scoperto da Aitchison e Brown (1957). Comunque, per completezza, esistono anche le formule che esprimono questi parametri in termini di p , n , m e sono:

$$\eta = \frac{2z}{\ln\left(\frac{m}{p}\right)}$$

$$\gamma^* = \eta \ln \left[\frac{\frac{m}{p} - 1}{p \sqrt{\frac{m}{p}}} \right]$$

$$\varepsilon = \frac{x_z - x_{-z}}{2} + \frac{p}{2} + \frac{\frac{m}{p} + 1}{\frac{m}{p} - 1}$$

4.3.4. Test per verificare la normalità dei dati trasformati

A questo punto, dopo aver definito la procedura di selezione della migliore trasformazione di Johnson da usare e la stima dei suoi parametri, bisogna determinare se i dati del campione trasformati sono dati distribuiti Normalmente. Questo viene svolto grazie al **test W di Shapiro-Wilk** e al **test di Kolmogorov-Smirnov** sui dati trasformati e, la trasformazione che ha il valore più elevato del test, è considerata la migliore.

4.3.4.1. Test di Shapiro-Wilk

La capacità di ognuna delle tre trasformazioni di Johnson di convertire i dati campionari in dati distribuiti Normalmente, è assicurata dal calcolo della statistica test **W** di Shapiro-Wilk per i dati trasformati la cui numerosità campionaria n è compresa tra 3 e 50. La procedura per calcolare la statistica test w è:

1. Disporre gli n dati in ordine crescente
2. Calcolare la somma dei quadrati SQ data da:

$$SQ = \sum_{i=1}^n (x_i - \bar{x})^2$$

dove \bar{x} è la media del campione.

3. Se n è pari, si pone $k = n/2$ e se n è dispari si pone $k = (n - 1)/2$, calcolando quindi il termine b come segue:

$$b = \sum_{i=1}^k a_{n-i+1} \cdot (x_{n-i+1} - x_i)$$

dove i valori di a sono ricavati da una tabella in funzione della numerosità campionaria.

4. Si calcola la statistica W pari a:

$$W = \frac{b^2}{SQ}$$

5. Infine si calcola il p-value per la statistica W e si hanno i seguenti casi:

- Se $p > 0.10$ allora la i dati trasformati sono distribuiti Normalmente e la trasformazione quindi è corretta
- Se $p < 0.10$ allora i dati trasformati sono distribuiti Non-Normalmente e la trasformazione non è corretta

Come è possibile immaginare, per calcolare la statistica test W e il suo p-value, esistono dei programmi statistici (come il FORTRAN per esempio) che facilitano di molto il calcolo.

4.3.4.2 Test di Kolmogorov-Smirnov.

Il test di Kolmogorov-Smirnov viene applicato ai dati trasformati dalla trasformazione di Johnson, dati la cui numerosità campionaria n è maggiore di 50.

Il metodo consiste nel determinare

1. Disporre i dati in ordine crescente

$$x_1 \leq x_2 \leq \dots \leq x_i \leq \dots \leq x_n \text{ con } i = 1, 2, \dots, n$$

2. La frequenza $F(x_i)$ reale e cumulata dei valori dei dati osservati dal campione ordinato, ossia:

$$F(x_i) = \frac{i}{n} \text{ dove } i = 1, 2, \dots, n$$

e il termine i rappresenta il valore ordinale dei singoli valori ordinati in ordine crescente.

3. Determinare gli scarti normalizzati, o scarti ridotti, dei singoli valori:

$$z_i = \frac{x_i - \bar{x}}{s}$$

dove x_i è l' i -esimo valore, \bar{x} la media aritmetica e s lo scarto quadratico medio

4. Dalla tavola della distribuzione normale cumulata di Gauss si ricavano i valori della probabilità $F(z_i)$ teorici della distribuzione stessa.

5. Si calcola lo stimatore di Kolmogorov-Smirnov D_2 pari a:

$$D_2 = \max_{i=1}^n [F(z_i) - F(x_i)]$$

6. Tale valori di D_2 trovato lo si confronta, ad un determinato livello di significatività α , con un valore di D , in funzione di n , riportato nella tavola dei valori percentili di D . Si ottiene quindi:
- a) Se $D_2 \leq D$ allora la distribuzione dei dati trasformati è da considerarsi Normale
 - b) Se $D_2 > D$ allora la distribuzione dei dati trasformati e da considerarsi Non - Normale

4.3.5. Classificazione delle più importanti distribuzioni secondo l'approccio di Johnson e relative conclusioni.

Chou (1994) ha presentato, in un suo articolo, alcune delle proprietà del rapporto RQ per alcuni valori di z per le distribuzioni più importanti: la uniforme, la Beta, l'Esponenziale, la Chi-Quadro, la Gamma, la F, la normale, la t di Student e la logistica e successivamente, basandosi sul criterio di Johnson e su quello di Slifker e Shapiro, queste distribuzioni sono state classificate attraverso l'analisi del loro range, degli indici (β_1, β_2) e del rapporto quantile RQ:

Distribuzioni	Classificazione attraverso il range	Classificazione attraverso (β_1, β_2)	Classificazione attraverso il RQ
Uniforme	S_B	S_B	S_B
Beta	S_B	S_B	S_B
Esponenziale	S_L	S_B	S_B
Chi-Quadro	S_L	S_B	S_B
Gamma	S_L	S_B	S_B
F	S_L	S_B, S_U, S_L	S_B, S_U, S_L
Normale	S_U	S_L	S_B
T di student	S_U	S_U	S_B
Logistica	S_U	S_U	S_B

Usando quindi i risultati presentati qui, è stato osservato come la classificazione basata su (β_1, β_2) e quella sul rapporto quantile RQ siano le più corrette per la maggioranza dei casi. A questo punto, Chou, dopo aver trovato questi risultati e fatto questa analisi, arriva alle seguenti conclusioni:

- a) Il rapporto RQ dipende dalla scelta di z . Questo significa che, per alcune distribuzioni come la F, il rapporto RQ potrebbe essere inferiore o superiore a 1, dipendendo tutto dalla scelta del valore che deve venire attribuito a z . Come risultato di questo quindi,

la scelta di z condiziona la classificazione fatta attraverso il sistema di Johnson, potendo quindi classificare una distribuzione limitata da un lato con S_B , S_U o S_L .

b) Le condizioni di Slifker e Shapiro del criterio da loro sviluppato, sono necessarie.

Questo significa che:

- Se una distribuzione limitata può essere trasformata in una distribuzione Normale standard, allora $RQ < 1$
- Se una distribuzione limitata da un lato può essere trasformata in una distribuzione Normale standard, allora $RQ = 1$
- Se una distribuzione non limitata può essere trasformata in una distribuzione Normale standard, allora $RQ > 1$

c) Le condizioni non sono sufficienti. Questo significa che:

- Se $RQ < 1$, allora la distribuzione potrebbe non essere limitata e quindi potrebbe non esistere una distribuzione S_B che trasformi la distribuzione in una distribuzione Normale.
- Se $RQ = 1$, allora la distribuzione potrebbe non essere limitata da un alto e quindi potrebbe non esistere una distribuzione S_L che trasformi la distribuzione in una distribuzione Normale.
- Se $RQ > 1$, allora la distribuzione potrebbe non essere non limitata e quindi potrebbe non esistere una distribuzione S_U che trasformi la distribuzione in una distribuzione Normale.

Applicando quindi il metodo dei percentili per adattare i dati distribuiti Non-Normalmente, sono state scoperte le seguenti limitazioni:

- a) Per una data serie di dati, l' RQ stimato dipendente dai percentili stimati è soggetto a variazioni campionarie.
- b) All'interno di un sistema di distribuzioni, si nota come alcune distribuzioni siano più difficili da adattare rispetto ad altre, non tutte le distribuzioni possono essere trasformate con una distribuzione di Johnson.
- c) Il problema di trovare la distribuzione del rapporto quantile stimato sembra essere intrattabile. Basandosi su delle simulazioni, la varianza di questo rapporto è abbastanza alta per campioni di elevata numerosità.
- d) Non sempre, adattando una serie di dati con una distribuzione di Johnson attraverso il criterio del rapporto RQ , si ottiene un adattamento adeguato. È necessario come visto fare dei test di normalità.
- e) Il metodo dei percentili non da una stima della variabilità o della distorsione degli indici stimati, non dando quindi un intervallo di confidenza per questi indici

4.4. Metodo di Clements, Metodo di Johnson-Kotz-Pearn e delle distribuzioni libere.

4.4.1. Introduzione.

Cominciamo con il dire che, modellare i dati con delle distribuzioni, è un metodo come visto molto appropriato nel caso di processi distribuiti Non-Normalmente. Il metodo di trasformare i dati con le distribuzioni o di Johnson o di Pearson è uno dei migliori, dal momento che queste sono capaci di adattare molte distribuzioni.

Questi metodi di Johnson e di Pearson sono conosciuti da circa cinquant'anni ormai e la loro partecipazione al processo di Controllo della Qualità è stata molto lenta. Sebbene quindi questa loro cosiddetta "vecchiaia", questi metodi non possono essere tralasciati da parte di chi ne ha bisogno, dal momento che recano dei rilevanti miglioramenti alle situazioni di Non-Normalità.

Come visto precedentemente dallo studio delle distribuzioni di Johnson, il metodo basato sulla stima dei quantili delle distribuzioni Non-Normali è molto usato. Questo metodo può essere anche usato per stimare direttamente gli Indici di Capacità dei processi così distribuiti. Infatti cominciamo considerando tale curva normale:

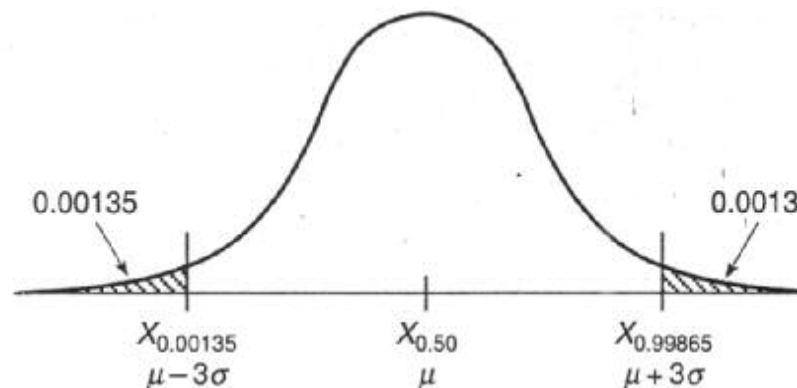


Figura 2

Si nota quindi come, per la distribuzione Normale, lo 0.135% dei dati sia sotto $\mu - 3\sigma$, il 50% dei dati sia sotto μ e il 99.865% dei dati sia sotto $\mu + 3\sigma$, lasciando quindi il 99.73% tra i due limiti $\mu \pm 3\sigma$. Possiamo definire i quantili $q(p)$ come :

$$\text{Pr ob}(X \leq q(p)) = p$$

e nel caso di una Normale standard si ha che

$$x_{0.135} = \mu - 3\sigma \quad ; \quad x_{0.50} = \mu \quad ; \quad x_{0.99865} = \mu + 3\sigma$$

ottenendo quindi che

$$\text{Pr ob}(X \leq x_{0.135}) = 0.135$$

$$\text{Pr ob}(X \leq x_{0.5}) = 0.5$$

$$\text{Pr ob}(X \leq x_{0.99865}) = 0.99865$$

Gli Indici di Capacità del processo possono, inoltre, essere ridefiniti nei termini dei loro quantili per ottenere così delle possibili modifiche nel caso di distribuzioni Non-Normali.

Questi Indici possono essere così definiti:

$$C_{p(q)} = \frac{USL - LSL}{x_{0.99865} - x_{0.00135}}$$

$$C_{pk(q)} = \min\left(\frac{USL - x_{0.5}}{x_{0.99865} - x_{0.5}}, \frac{x_{0.5} - LSL}{x_{0.5} - x_{0.00135}}\right)$$

Nel caso di processi Non-Normali quindi, se si riconoscesse quale distribuzione segue il processo, si potrebbe adattare meglio i nostri dati con delle trasformazioni (verificando sempre con un test il buon adattamento) e ottenere così delle accurate misure dei quantili del processo.

La seguente figura infatti

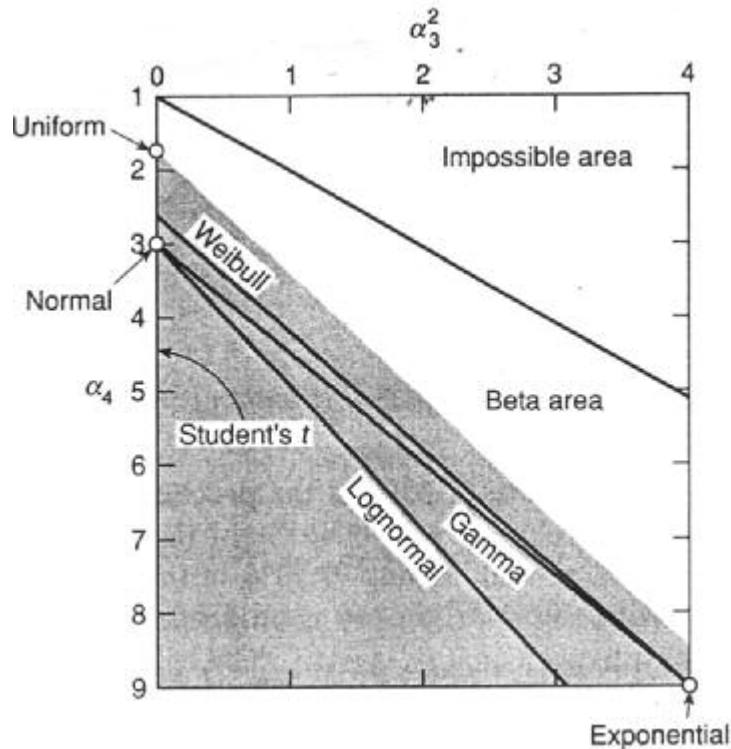


Figura 3

è un grafico del piano (β_1, β_2) associato ad alcune distribuzioni. Se il vero valore, o quello stimato, dell'asimmetria e della curtosi derivante dai dati del processo fosse conosciuto, questa figura potrebbe allora essere usata per scegliere il miglior modello per i nostri dati. Come detto però, solo un test di buon adattamento potrebbe assicurarci la scelta, ma questo grafico potrebbe comunque essere di buon aiuto. Come segue quindi, il metodo più conosciuto per la trasformazione quantile dei dati è quello che è stato sviluppato da Clements (1989), studioso che ha usato le curve di Pearson per ottenere le stime migliori dei tre quantili.

4.4.2. Il metodo di Clements.

Clements (1989) propose un metodo per la correzione del C_p e del C_{pk} basato sull'assunzione che la distribuzione può essere adeguatamente rappresentata da una distribuzione di Pearson.

In pratica si deve sostituire nel calcolo di C_{pk} il moltiplicatore 6, al denominatore, con il parametro θ , ottenendo così

$$C_{pk} = \min\left(\frac{\mu - LSL}{\theta \cdot \sigma}, \frac{USL - \mu}{\theta \cdot \sigma}\right)$$

tale che la probabilità sia:

$$Prob\left[\mu - \frac{1}{2} \theta \cdot \sigma \leq x \leq \mu + \frac{1}{2} \theta \cdot \sigma\right] = 0.0027$$

Per determinati valori dell'asimmetria ($\sqrt{\beta_1}$) e della curtosi (β_2), dalla tabella sottostante si ricavano i parametri θ_u e θ_l tale che

$$Prob[x \leq \mu - \theta_l \cdot \sigma] = 0.135\% \quad e \quad Prob[x \leq \mu + \theta_u \cdot \sigma] = 99.685\%$$

ricavando così il parametro θ attraverso la relazione

$$\theta = \theta_u - \theta_l$$

La tabella rappresenta i punti standardizzati a 0.135% e 99.865% della curva di Pearson con $\sqrt{\beta_1} > 0$. Se $\sqrt{\beta_1} < 0$ allora bisogna scambiare 0.135% e 99.865% con il cambio di segno.

$\sqrt{\beta_1} \backslash \beta_2$	0.0	0.2	0.4	0.6	0.8	1.0	1.2	1.4	1.6	1.8
1.8	-1.727 1.727	-1.496 1.871	-1.230 1.896	-0.975 1.803	-0.747 1.636	- -	- -	- -	- -	- -
2.2	-2.210 2.210	-1.912 2.400	-1.555 2.454	-1.212 2.349	-0.927 2.108	-0.692 1.822	- -	- -	- -	- -
2.6	-3.000 3.000	-2.535 2.869	-1.930 2.969	-1.496 2.926	-1.125 2.699	-0.841 2.314	-0.616 1.928	- -	- -	- -
3.0	-3.000 3.000	-2.689 3.224	-2.289 3.358	-1.817 3.385	-1.356 3.259	-1.000 2.914	-0.739 2.405	-0.531 1.960	- -	- -
3.4	-3.261 3.261	-2.952 3.484	-2.589 3.639	-2.127 3.175	-1.619 3.681	-1.178 3.468	-0.865 2.993	-0.634 2.398	- -	- -
3.8	-3.458 3.458	-3.118 3.678	-2.821 3.844	-2.396 3.951	-1.887 3.981	-1.381 3.883	-1.000 3.861	-0.736 2.945	-0.533 2.322	- -
4.2	-3.611 3.611	-3.218 3.724	-2.983 3.997	-2.616 4.124	-2.132 4.194	-1.602 4.177	-1.149 3.496	-0.840 3.529	-0.617 2.798	- -
4.6	-3.731 3.371	-3.282 3.942	-3.092 4.115	-2.787 4.253	-2.345 4.351	-1.821 4.386	-1.316 4.311	-0.950 4.015	-0.701 3.364	-0.510 2.609
5.0	-3.828 3.828	-3.325 4.034	-3.167 4.208	-2.914 4.354	-2.524 4.468	-2.023 4.539	-1.494 4.532	-1.068 4.372	-0.785 3.907	-0.580 3.095

Nel calcolo del C_{pk} Clements suggerisce il medesimo valore di θ ma con la sostituzione della media con la mediana M .

4.4.2.1. Uso delle curva di distribuzione di Pearson.

Clements ha quindi proposto un metodo per modificare il C_p e il C_{pk} per le distribuzioni Non-Normali, usando la curva di Pearson per ottenere un'accurata stima dei quantili $x_{0.00135}$, $x_{0.5}$ e $x_{0.99865}$.

Questo metodo si sviluppa seguendo i seguenti passi:

- a) Si stima la media campionaria \bar{x} , la deviazione standard s , l'asimmetria e la curtosi dei dati provenienti dal processo. Inoltre si devono conoscere il limite superiore USL e quello inferiore LSL. È da notare come Clements usi la misura standardizzata dell'asimmetria (β_1) e della curtosi ($\beta_2 - 3$)
- b) Basandosi sulle stime dell'asimmetria e della curtosi, si usano due tavole per determinare L'_p e U'_p , che rispettivamente sono il percentile 0.135% e il percentile 99.865% della distribuzione. Queste due tavole sono in funzione dell'asimmetria e della curtosi e vanno usate in una maniera particolare, infatti: in una si trova il valore di L'_p per asimmetria positiva e il valore di U'_p per asimmetria negativa e nell'altra il valore di L'_p per asimmetria negativa e il valore di U'_p per asimmetria positiva.
- c) Si calcola M , la mediana, il percentile 50%, usando una tavola che è sempre in funzione dell'asimmetria e della curtosi. Il valore che si trova all'interno di questa tavola deve essere:
 - cambiato di segno se l'asimmetria è positiva
 - lasciato uguale se l'asimmetria è negativa

d) Si stima $x_{0.00135}$:

$$L_p = \bar{x} - L'_p \cdot s$$

e) Si stima $x_{0.5}$:

$$M = \bar{x} + M' \cdot s$$

f) Si stima $x_{0.99865}$:

$$U_p = \bar{x} + U'_p \cdot s$$

g) Si stima infine gli Indici di Capacità del processo:

- $\hat{C}_p = \frac{USL - LSL}{U_p - L_p}$
- $\hat{C}_{pl} = \frac{M - LSL}{M - L_p}$ e $\hat{C}_{pu} = \frac{USL - M}{U_p - M}$ ottenendo quindi che $\hat{C}_{pk} = \min(\hat{C}_{pl}; \hat{C}_{pu})$

$$\begin{aligned} \bullet \hat{C}_{pm} &= \frac{USL - LSL}{6 \sqrt{\left[\left(\frac{s(U_p' - L_p')}{6} \right)^2 + (\bar{x} + (M' \cdot s) - T)^2 \right]}} \\ \bullet \hat{C}_{pm}^* &= \min \left[\frac{USL - T}{3 \sqrt{\left[\left(\frac{s(U_p' - L_p')}{6} \right)^2 + (\bar{x} + (M' \cdot s) - T)^2 \right]}} \right. \\ &\quad \left. \frac{T - LSL}{3 \sqrt{\left[\left(\frac{s(U_p' - L_p')}{6} \right)^2 + (\bar{x} + (M' \cdot s) - T)^2 \right]}} \right] \\ \bullet \hat{C}_{pmk} &= \min \left[\frac{USL - (\bar{x} + M' \cdot s)}{3 \sqrt{\left[\left(\frac{s(U_p' - L_p')}{3} \right)^2 + (\bar{x} + (M' \cdot s) - T)^2 \right]}} \right. \\ &\quad \left. \frac{(\bar{x} + M' \cdot s) - LSL}{3 \sqrt{\left[\left(\frac{s(U_p' - L_p')}{3} \right)^2 + (\bar{x} + (M' \cdot s) - T)^2 \right]}} \right] \end{aligned}$$

Bisogna quindi fare alcune annotazioni fondamentali. Focalizzandoci quindi su i due Indici di Capacità di base C_p e C_{pk} , la prima annotazione riguarda il perché si deve usare la mediana per misurare la centralità del processo. Il motivo di questo risiede nell'assicurare che la misura di C_{pu} e C_{pl} misuri la relazione tra le parti superiori ed inferiori della metà della distribuzione dei dati e le tolleranze inferiori e superiori. La seconda annotazione è che le modifiche di Clements fatte per C_p e C_{pk} sono molto apprezzate dal momento che non richiedono trasformazioni dei dati e sono facili da trovare anche semplicemente con un calcolatore. È stato però proposto sempre da Clements l'uso della classica stima per la curtosi, precisamente

$$\beta_2 = \frac{m_4}{s^4} - 3 \quad \text{con} \quad m_4 = \frac{\sum_{i=1}^n (x_i - \bar{x})^4}{n}$$

Infine, ultima annotazione, riguarda la classifica dei quattro Indici C_p , C_{pk} , C_{pm} e C_{pmk} , facendo però attenzione alla sensibilità con cui la mediana del processo devia dal valore obiettivo. La classifica è identica a quella del caso Normale, quindi:

1. C_{pmk}
2. C_{pm}
3. C_{pk}
4. C_p

4.4.3. Il metodo di Johnson-Kotz-Pearn.

L'applicazione del metodo di Clements richiede la conoscenza dei coefficienti $\sqrt{\beta_1}$ e β_2 , che non sempre è facile ottenere. Quindi sono stati cercati degli approcci differenti per cercare di superare questa difficoltà, e uno di questi è quello proposto da Johnson (1992), che ha definito

$$C_{p(\theta)} = \frac{USL - LSL}{\theta \cdot \sigma} = \frac{d}{\frac{1}{2} \theta \cdot \sigma}$$

con θ che è una costante scelta in modo tale che la probabilità

$$\text{Prob} \left[\mu - \frac{1}{2} \cdot \sigma \leq x \leq \mu + \frac{1}{2} \cdot \sigma \right] = p$$

sia minore di uno e indipendente quanto possibile dalla distribuzione di X.

La scelta migliore per θ può provenire dai risultati trovati da Pearson e Tukey (1965), i quali hanno esaminato, per diverse distribuzioni Chi-Quadro, i valori di θ tali che la miglior percentuale di prodotti conformi abbia un valore specifico, ottenendo la seguente tabella:

Gradi di libertà	6	8	10	12	15	20	30	60	∞
β_1	1.33	1.00	0.80	0.67	0.53	0.40	0.27	0.13	0.00
β_2	5.00	4.50	4.20	4.00	3.80	3.60	3.40	3.20	3.00
θ per $P = 0.95$	3.82	3.84	3.85	3.87	3.88	3.89	3.90	3.91	3.92
θ per $P = 0.98$	4.60	4.61	4.62	4.62	4.63	4.63	4.64	4.64	4.65
θ per $P = 0.99$	5.16	5.15	5.15	5.15	5.15	5.15	5.15	5.15	5.15

Si può notare come $\theta = 5.15$ dia dei valori molto stabili per la percentuale di prodotti non conformi, circa l'1%.

Pearn raccomanda l'uso di $\theta = 5.15$, ottenendo così

$$C_{p(5.15)} = \frac{2d}{5.15 \cdot \sigma} = \frac{d}{2.575 \cdot \sigma}$$

e osservando tuttavia come questa stabilità sia raggiunta attraverso la sostituzione del 27% dei prodotti non conformi per gli Indici distribuiti Normalmente con l'1% dei prodotti non conformi.

Per quanto riguarda il C_{pk} invece, la corrispondente variazione è:

$$C_{pk(5.15)} = \frac{d - \left| \mu - \frac{1}{2} (LSL + USL) \right|}{2.575 \cdot \sigma}$$

ricordando che sostituendo σ con s si può trovare una stima sia per $C_{pk(5.15)}$ che per $C_{p(5.15)}$.

4.4.3.1. Differenze tra il metodo di Clements e quello di Johnson-Kotz-Pearn.

Le differenze principali tra il metodo di Clements e quello di Johnson-Kotz-Pearn sono le seguenti:

- a) Nel metodo di Clements c'è un tentativo di assegnare direttamente i valori dell'asimmetria e della curtosi, mentre nel secondo metodo si cerca di dare dei limiti che non sono sensibili a questi due valori.
- b) Nel secondo metodo non si garantisce a lunga scadenza una uguale probabilità limite, ma non si devono comunque stimare $\sqrt{\beta_1}$ e β_2 che sono difficili da ottenere con precisione dal momento che sono soggetti a grande variazione.
- c) Entrambi i metodi fanno affidamento sull'assunzione che la distribuzione della popolazione deve avere un modello unimodale, per il metodo di Clements vicino a quello della distribuzione di Pearson, mentre per il metodo di Johnson-Kotz-Pearn vicino a quello delle distribuzioni Gamma.

4.4.3.2. L'approccio di Munechika.

Munechika (1986) ha cercato di utilizzare un rapporto approssimativo tra il corrispondente percentile della Normale standard e la distribuzione Gram-Charlier (Edgeworth), cercando di ottenere una relazione approssimativa tra il valore dell'Indice di Capacità del processo distribuito Non-Normalmente e il rispettivo valore corretto, applicando dunque questo metodo solamente nel caso in cui ci sia il limite di specifica superiore USL.

Munechika è stato quindi uno dei primi ad utilizzare una modifica di C_{pk} per processi distribuiti Non-Normalmente valendosi dell'espansione di Cornish-Fisher. Espansione che è essenzialmente un'estensione delle distribuzioni continue attraverso una serie infinita di termini di una variabile casuale distribuita Normalmente, che può essere usata per sostituire il 6σ al denominatore di C_{pk} nel caso Normale con il $X_{0.99865} - X_{0.00135}$ nel caso Non-Normale. Ricordiamo che una variabile casuale X standardizzata con media 0 e varianza 1 può essere espressa come

$$X = U + B_1(U) + B_2(U) + \dots + B_k(U)$$

con dei particolari valori di $B_1(U)$ e $B_2(U)$.

È stato dimostrato quindi come, troncando l'espansione dopo $B_1(U)$ ed immaginando che X sia una Chi-Quadrato con n gradi di libertà, questo errore nel troncamento sia trascurabile.

Questo ha portato a correggere il C_{pk} con $C_{p\lambda}$, definendolo come

$$C_{p\lambda} = \lambda C_{pk}$$

dove λ è un coefficiente di correzione definito come

$$\lambda = \frac{\sqrt{\sqrt{\beta_1}^2 + 18 \cdot C_{pk} \cdot \sqrt{\beta_1} + 9} - 3}{3 \cdot C_{pk} \cdot \sqrt{\beta_1}}$$

dove k_3 è il terzo momento centrale

Quindi, dopo aver dato queste definizioni, posso procedere con la definizione della procedura di Munechika:

1. Si calcola $\bar{x} = \frac{\sum x_i}{n}$, $s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$ e $\sqrt{\hat{\beta}_1} = \frac{\sqrt{n} \sum (x_i - \bar{x})^3}{[\sum (x_i - \bar{x})^2]^{3/2}}$.
2. Si calcola $\hat{C}_{pk} = \frac{(USL - \bar{x})}{3s}$.
3. Si calcola $\hat{\lambda}$ sostituendo nella formula precedente a C_{pk} e a k_3 la loro stima.
4. Si calcola $C_{p\lambda} = \lambda C_{pk}$

4.4.4. Indici di Capacità per le distribuzioni 'libere'.

Questo metodo delle distribuzioni 'libere' è stato proposto da Chan (1988), il quale ha infatti scoperto come la stima della varianza s presente al denominatore di \hat{C}_p sia solamente usata per stimare la lunghezza (6σ) dell'intervallo di confidenza, intervallo che racchiude il 99.73% dei valori di X (sempre sotto l'assunzione che i dati siano distribuiti Normalmente e perfettamente centrati nel centro dell'intervallo, in $\mu = \frac{1}{2}(USL + LSL)$).

Chan ha quindi proposto, per stimare la lunghezza di questi intervalli, l'uso degli intervalli di confidenza per le distribuzioni libere. Questi intervalli sono stati costruiti per contenere al minimo, con una probabilità pari al $100 \cdot (1-\alpha)\%$, il $100 \cdot \beta\%$ della distribuzione per un dato valore di α (di solito vicino a 0) e di β (di solito vicino a 1).

Solitamente la scelta per β cade su $\beta = 0.9973$ e quindi su $\alpha = 0.05$, ma sfortunatamente questo richiede un campione di numerosità campionaria superiore a 1000. Quindi, per superare questa difficoltà, è stato osservato come, costruendo intervalli di confidenza con piccoli valori di β ma sempre con $\alpha = 0.05$, questo richieda campioni di numerosità campionaria inferiore a 300. È stato sottolineato come:

- a) Con $\beta = 0.9546$ si usi $\frac{3}{2}s$ al posto di $6s$
- b) Con $\beta = 0.6826$ si usi $3s$ al posto di $6s$

Le basi per la scelta sono quelle usati nel caso di distribuzioni normali:

- a) L'intervallo $(\mu - 2\sigma; \mu + 2\sigma)$ di lunghezza 4σ contiene il 95.46% dei dati
- b) L'intervallo $(\mu - \sigma; \mu + \sigma)$ di lunghezza 2σ contiene il 68.26% dei dati

4.5. Altri fondamentali Indici di Capacità per le distribuzioni Non-Normali e Asimmetriche.

4.5.1. Un Indice di Capacità flessibile: $C_{j_{kp}}$.

Questo Indice venne introdotto da Johnson nel 1994 con la precisa funzione di tenere conto delle possibili differenze tra la variabilità di X per valori sopra e sotto il valore obiettivo T. $C_{j_{kp}}$ ha il vantaggio, rispetto agli altri Indici, di cambiare velocemente il suo valore quando la distribuzione del processo cambia.

Vengono definiti per prima cosa $CU_{j_{kp}}$ e $CL_{j_{kp}}$, rispettivamente la Capacità della metà superiore ed inferiore dei dati del processo:

$$CU_{j_{kp}} = \frac{1}{3\sqrt{2}} \left[\frac{(USL - T)}{\sqrt{E_{X>T} [(X - T)^2]}} \right] \text{ e } CL_{j_{kp}} = \frac{1}{3\sqrt{2}} \left[\frac{(T - LSL)}{\sqrt{E_{X\leq T} [(X - T)^2]}} \right]$$

dove al denominatore, ho le "semivarianze"

$$E_{X<T} [(X - T)^2] = E[(X - T)^2 | X < T] \cdot \text{Prob}(X < T)$$

e

$$E_{X>T} [(X - T)^2] = E[(X - T)^2 | X > T] \cdot \text{Prob}(X > T)$$

Da notare come i fattori $\text{Prob}(X > T)$ e $\text{Prob}(X < T)$ tengano conto di quanta deviazione positiva o negativa ci sia dal valore obiettivo T.

Quindi, dopo aver definito $CU_{j_{kp}}$ e $CL_{j_{kp}}$, si definisce $C_{j_{kp}}$ come

$$C_{j_{kp}} = \min(CU_{j_{kp}}, CL_{j_{kp}})$$

$$C_{j_{kp}} = \frac{1}{3\sqrt{2}} \min \left[\frac{USL - T}{\sqrt{E_{X>T} [(X - T)^2]}}, \frac{T - LSL}{\sqrt{E_{X\leq T} [(X - T)^2]}} \right]$$

Uno stimatore naturale di $C_{j_{kp}}$ è

$$\hat{C}_{j_{kp}} = \frac{1}{2\sqrt{3}} \min \left(\frac{USL - T}{\sqrt{\frac{s_+}{n}}}, \frac{T - LSL}{\sqrt{\frac{s_-}{n}}} \right)$$

dove

$$s_+ = \sum_{X_i > T} (x_i - T)^2 \text{ e } s_- = \sum_{X_i < T} (x_i - T)^2$$

e da uno studio di Kotz e Johnson, si rileva dalla tabella contenente i valori di $E(\hat{C}_{j_{kp}} / C_{j_{kp}})$ e $D.S.(\hat{C}_{j_{kp}} / C_{j_{kp}})$ come la stima di $C_{j_{kp}}$ in generale sia distorta. Questa distorsione è negativa quando il valore obiettivo $T = \frac{1}{2}(USL + LSL)$ ma aumenta quando $(USL - T)/d$

decrese, aumento che diventa sempre più grande quando il valore di T è circa attorno a 0, 4. La varianza di questa stima decresce all'aumentare della numerosità campionaria n , infatti aumenta quando $(USL - T)/d$ diminuisce e quando il valore obiettivo si avvicina al limite di specifica superiore. Anche la distorsione aumenta con l'aumentare di n , effetto che è particolarmente notevole per piccoli valori di $(USL - T)/d$.

Franklin e Wasserman (1994) hanno successivamente cercato di determinare il limite di specifica inferiore al 95% per un campione di numerosità campionaria abbastanza grande, per $n = 1000$. Il metodo da loro usato è stato quello della simulazione, metodo che ha procurato un accettabile copertura per il limite di specifica inferiore al 95% nel caso in cui $\mu = T$, copertura percentuale che è stata tabulata usando una distribuzione simmetrica, la Normale e una non simmetrica, la Chi-Quadro. Se, per qualsiasi motivo, i dati del processo tendono a diventare asimmetrici, quindi $\mu \neq T$, l'Indice C_{jpk} rimane il più robusto nel rispettare le performance rispetto agli altri Indici C_p , C_{pk} e C_{pm} .

4.5.2. Indice C_s di Wright.

Wright (1995) è stato il promotore dell'indice di Capacità C_s , indice che è stato ricavato da una modifica di C_{pmk} per il caso Non-Normale e unimodale, dovuta fondamentalmente all'aggiunta di un fattore di asimmetria all'interno del suo denominatore. Quindi, con questa aggiunta, si può notare come questo Indice sia specialmente adatto per quei processi dove l'asimmetria è una importante componente del peggioramento della Capacità, come per esempio i processi di perforazione e i processi caratterizzati da caratteristiche qualitative quali, l'essere circolare, cilindrico, diritto e piano, aspetti dove l'asimmetria positiva è di norma. Infatti la distribuzione dei diametri di un foro, prodotti da un processo di perforazione in controllo statistico, è solitamente vicino ad una distribuzione normale troncata a sinistra e con una asimmetria a destra. La caratteristica fondamentale di processo di perforazione che sta peggiorando, si manifesta con un simultaneo aumento della media, della varianza e della asimmetria e con un piccolo cambiamento nei percentili inferiori della distribuzione.

C_s è stato originariamente presentato per essere quindi sensibile all'asimmetria e per diminuire quando questa aumenta all'interno della distribuzione presa in considerazione, vedendo quindi l'asimmetria come qualcosa di dannoso. In un secondo momento, si guarda a C_s con un'altra prospettiva, sottolineando come questo possa concedere la possibilità di essere sfruttato per determinare le performance dei processi distribuiti Non-Normalmente, mantenendo le caratteristiche di applicabilità di C_{pmk} per quei processi che possono avere la media non centrata all'interno dei limiti di specifica e che possono non essere centrati nel valore obiettivo.

Quindi C_s viene definito nel modo seguente:

$$C_s = \frac{\min(USL - \mu, LSL - \mu)}{3\sqrt{\sigma^2 + (\mu - T)^2 + |\mu_3/\sigma|}} = \frac{d - |\mu - T|}{3\sqrt{\sigma^2 + (\mu - T)^2 + |\mu_3/\sigma|}}$$

con $d = (USL - LSL)/2$, dove il momento terzo $\mu_3 = E(X - \mu)^3$ è diviso per σ , assicurando così che il termine di asimmetria sia espresso nella stessa unità degli altri termini del denominatore, e dove $|\mu_3/\sigma|$ assicura che una asimmetria negativa penalizzerebbe l'Indice riducendolo.

Questo Indice quindi è usato per indicare quanta sensibilità ci sia verso i cambiamenti che si sviluppano all'interno delle distribuzioni considerate.

Se tutti gli elementi di C_s vengono divisi per σ , si ottiene

$$C_s = \frac{d/\sigma - |(\mu - T)/\sigma|}{3\sqrt{1 + ((\mu - T)/\sigma)^2 + |\sqrt{\beta_1}|}}$$

dove $\sqrt{\beta_1}$ è il fattore standardizzato dell'asimmetria e dove, come per gli Indici di Capacità tradizionali, un valore di C_s pari a 1 viene considerato il limite inferiore perché il processo venga considerato Capace.

Da notare anche come $C_s < C_{pmk}$ e come l'uguaglianza tra questi due sia raggiunta solo nel caso di un processo simmetrico. Come per gli Indici di Capacità tradizionali, un valore di $C_s = 1$ è considerato la soglia minore.

Una logica modifica di C_s potrebbe essere quella di cambiare l'effetto causato sull'Indice dal parametro di asimmetria μ_3/σ , omettendo il suo valore assoluto e sostituendolo con il suo valore reciproco $-\mu_3/\sigma$, ottenendo così

$$C_{sm} = \frac{\min(USL - \mu, LSL - \mu)}{3\sqrt{\sigma^2 + (\mu - T)^2 - \mu_3/\sigma}} = \frac{d - |\mu - T|}{3\sqrt{\sigma^2 + (\mu - T)^2 - \mu_3/\sigma}}$$

Con questa nuova definizione, un valore positivo dell'asimmetria avrebbe l'effetto di aumentare il valore dell'Indice, mentre un valore negativo dell'asimmetria avrebbe l'effetto di diminuire il valore dell'Indice.

4.5.2.1. Lo stimatore di C_s .

Uno stimatore naturale di C_s è

$$\hat{C}_s = \frac{USL - \bar{x}}{3\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - T)^2 + \left| \frac{n^2 m_3}{(n-1)} \div \sqrt{\frac{n}{n-1} \cdot \frac{m_2}{c_4^2}} \right|}}$$

dove $m_r = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^r$ è l'r-esimo momento centrale, ottenendo quindi

$$m_2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{e} \quad m_3 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3$$

e c_4 è pari a

$$c_4 = \sqrt{\frac{2}{n-1} \cdot \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}} \cong \frac{4(n-1)}{4n-3}$$

Per quanto riguarda C_{sm} , un suo stimatore naturale è:

$$\hat{C}_{sm} = \frac{(USL - \bar{x})}{3 \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - T)^2 + \frac{n \cdot c_4}{\sqrt{(n-1)(n-2)}} \sum_{i=1}^n (x_i - \bar{x})^3 + \sum_{i=1}^n (x_i - \bar{x})^2}}$$

Valutando quindi la distorsione percentuale all'interno di C_s e C_{pmk} ($\text{dist.C} = [E(C) - C]/C$), Wright ha trovato che \hat{C}_s è più precisa di \hat{C}_{pmk} per piccoli campioni derivanti da processi che non raggiungono il valore obiettivo, dove appaiono quindi significativi salti dalla media.

Infatti quando $\mu = T$, \hat{C}_s è più distorta rispetto a \hat{C}_{pmk} , variando le distorsioni rispettivamente tra -11% e -21% e tra il -5% e 4%. Quando invece $\mu \neq T$, le cose cambiano. Per campioni elevati \hat{C}_{pmk} tende ad avere minore distorsione rispetto a \hat{C}_s , sotto la condizione però di piccoli salti dalla media e per d/σ elevato. Quando invece le riduzioni della Capacità diventano più critiche a causa di elevati salti dalla media e d/σ piccolo, \hat{C}_s tende ad avere minore distorsione rispetto a \hat{C}_{pmk} . Nessuno dei due Indici dimostra una distorsione maggiore del $\pm 10\%$ quando il processo non è centrato nel valore obiettivo, distorsione che si riduce con l'aumentare dei salti dalla media.

Chen e Johnson (1996) hanno ricavato le proprietà asintotiche della distribuzione di C_s sotto quattro condizioni: (1) $\mu = T$ e $\sqrt{\beta_1} = 0$, (2) $\mu = T$ e $\sqrt{\beta_1} \neq 0$, (3) $\mu \neq T$ e $\sqrt{\beta_1} = 0$, (4) $\mu \neq T$ e $\sqrt{\beta_1} \neq 0$ ed hanno trovato che \hat{C}_s tende a C_s , essendo quindi consistente ed asintoticamente non distorto per tutti e quattro i casi. In più hanno scoperto che il comportamento asintotico di \hat{C}_s è significativamente sensibile all'asimmetria ed hanno approvato la conclusione fatta da Wright, che la distribuzione asintotica della stima di C_s è normale quando $\mu \neq T$ e $\sqrt{\beta_1} \neq 0$. Questo ha portato a fare di questa una naturale scelta per lo studio della Capacità di un processo per campioni di elevata numerosità campionaria, senza guardare alla distribuzione dei dati. In generale il C_s fornisce un approccio statistico per l'analisi della capacità del processo sotto tre condizioni:

- $USL - \mu \neq \mu - LSL$ (per esempio $\mu \neq M = \frac{1}{2}(USL + LSL)$)

- $\mu \neq T$
- $\sqrt{\beta_1} \neq 0$.

Comunque è sorta una discussione su quale sia il significato pratico di C_s . Vannman (1995) ha affermato che è inappropriato il punto di vista di chi pensa che l'asimmetria sia qualcosa di negativo, dovuta soprattutto alla difficoltà di molti pratici di ricavare C_s . La cosa più importante è che non ci sono legami diretti con la percentuale NC di prodotti non conformi, lasciando quindi i pratici stupiti su quale sia il significato della stima di C_s . Infatti, nel caso degli Indici di prima e seconda generazione, c'era un legame tra la loro stima e la percentuale NC, legame che sottolinea come ad una elevato Indice di Capacità sia associata una bassa percentuale di prodotti non conformi e che nel caso di C_s non appare. Dal momento che non è possibile trovare una relazione nella forma di limite tra la percentuale NC e l'Indice, non è possibile fare un confronto diretto per questi Indici.

È per questo che quindi viene definito l'Indice modificato C_{sm} , Indice che cerca di riflettere la visione dell'asimmetria come 'qualcosa di buono', suggerendo che un Indice può essere valutato per quanto sensibile è ai cambiamenti che si verificano all'interno della distribuzione presa in considerazione. Non c'è nessun riferimento al caso Normale, ma grazie a C_{sm} , per quei pratici che lavorano con serie di dati asimmetrici, è quindi possibili cambiare prospettiva. Grazie alla differente prospettiva c'è la possibilità che un Indice, che originariamente era usato per sottolineare come fosse qualcosa di penalizzante l'asimmetria nelle distribuzioni, venga usato con una interpretazione e forma diversa.

4.5.3. Indici di Capacità basati sul metodo della deviazione standard pesata (WSD).

4.5.3.1. Il metodo della deviazione standard pesata (WSD).

Bai e Choi (2001) hanno proposto un nuovo metodo euristico per costruire semplici Indici di Capacità per le popolazioni asimmetriche, metodo che si basa sul metodo della deviazione standard (WSD). Questo metodo corregge i valori degli Indici di Capacità di base, in accordo con i gradi di asimmetria della distribuzione presa in considerazione ed usando differenti fattori nel calcolo della deviazione sopra e sotto la media del processo. Quando la distribuzione considerata è simmetrica, questi Indici si riducono a quelli di base.

Chan e Bai hanno quindi inizialmente proposto questo metodo della WSD per costruire carte di controllo in accordo con i gradi di asimmetria delle distribuzioni, basandosi sull'idea che σ_x può essere diviso nella deviazione superiore ed inferiore, σ_U^W e σ_L^W , deviazioni che rappresentano rispettivamente i gradi di dispersione del lato sopra e sotto μ .

Una funzione di densità asimmetrica $f(x)$ può essere approssimata con due funzioni di densità Normali,

$$f_U(x) = \frac{1}{2\sigma_U^W} \Phi\left(\frac{x - \mu_x}{2\sigma_U^W}\right) \quad \text{e} \quad f_L(x) = \frac{1}{2\sigma_L^W} \Phi\left(\frac{x - \mu_x}{2\sigma_L^W}\right)$$

che hanno la stessa media μ_x ma differenti deviazioni standard $2\sigma_U^W$ e $2\sigma_L^W$, come viene rappresentato nella figura 4, dove $\Phi(\cdot)$ è la funzione di densità della Normale standard.

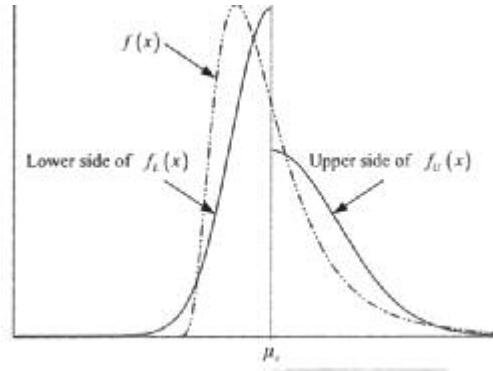


Figura 4

Per esempio i lati superiori ed inferiori di $f(x)$ sono approssimati rispettivamente con il lato superiore di $f_U(x)$ e il lato inferiore di $f_L(x)$.

Le deviazioni standard σ_U^W e σ_L^W sono definite quindi come:

$$\sigma_U^W = 2P_x \cdot \sigma_x \quad \text{e} \quad \sigma_L^W = 2(1 - P_x) \cdot \sigma_x$$

dove $P_x = \text{Prob}(X \leq \mu_x)$

La stima di queste deviazioni è:

$$s_U^W = 2P_x \cdot s_x \quad \text{e} \quad s_L^W = 2(1 - P_x) \cdot s_x$$

dove $P_x = \text{Prob}(X \leq \bar{x})$ mentre $s_x = \sum(x_i - \bar{x})^2/n-1$.

Vengono anche definite le deviazioni standard dal valore obiettivo sempre basandosi sull'idea che σ_x può essere diviso nella deviazione superiore ed inferiore. Queste deviazioni, denominate rispettivamente σ_{T1}^W e σ_{T2}^W , vengono usate per la correzione degli Indici di seconda e terza generazione.

La loro definizione è:

$$\sigma_{T1}^W = 2P_T \cdot \sigma_T \quad \text{e} \quad \sigma_{T2}^W = 2(1 - P_T) \cdot \sigma_T$$

dove $P_T = \text{Prob}(X \leq T)$

La stima di queste deviazioni è:

$$s_{T1}^W = 2P_T \cdot s_T \quad \text{e} \quad s_{T2}^W = 2(1 - P_T) \cdot s_T$$

dove $P_T = \text{Prob}(X \leq T)$ mentre $s_T = \sum(x_i - T)^2/n-1$.

4.5.3.2. L'Indice C_p basato sul metodo WSD.

Il C_p basato sul metodo WSD è denominato C_p^{WSD} ed definito come:

$$C_p^{WSD} = \min \left\{ \frac{USL - LSL}{6 \cdot 2\sigma_U^W}, \frac{USL - LSL}{6 \cdot 2\sigma_L^W} \right\} = \min \left\{ \frac{USL - LSL}{6 \cdot 2P_x \sigma_x}, \frac{USL - LSL}{6 \cdot 2(1 - P_x) \sigma_x} \right\}$$

$$= \frac{USL - LSL}{6\sigma_x} \min \left\{ \frac{1}{2P_x}, \frac{1}{2(1 - P_x)} \right\}$$

$$= \frac{C_p}{D_x}$$

dove $D_x = 1 + |1 - 2P_x|$ e dove le deviazioni σ_U^W e σ_L^W sono usate al posto di σ_x per riflettere i gradi di asimmetria. Se la distribuzione presa in considerazione è simmetrica, allora $P_x = 0.5$ e $D_x = 1.0$ cosicché $C_p^{WSD} = C_p$. Se invece la distribuzione è asimmetrica, allora $D_x > 1.0$ e $C_p^{WSD} < C_p$.

Per fare un esempio, la seguente tabella presenta gli effetti di C_p^{WSD} per un processo A distribuito come una $N(40, 10^2)$, per B distribuito come una $LN(3.39, 0.31, 8.67)$, per C distribuito come una $LN(2.68, 0.55, 23.23)$ e per D distribuito come una $LN(2.24, 0.72, 29.31)$

Processo	$\sqrt{\beta_1}$	P_x	PNC	C_p^{WSD}	C_p
A	0.00	0.50	2700	1.00	1.00
B	1.00	0.56	10461	0.89	1.00
C	2.00	0.61	16358	0.82	1.00
D	3.00	0.64	18325	0.78	1.00

La funzione di densità della distribuzione lognormale è

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_i(x - \theta_i)} \exp \left[-\frac{\{\ln(x - \theta_i) - \zeta_i\}^2}{2\sigma_i^2} \right]$$

dove i parametri $(\zeta_i, \sigma_i, \theta_i)$ sono determinati in modo che i processi B, C e D abbiano la stessa media $\mu_x = 40$ e deviazione standard $\sigma_x = 10$ ma differente asimmetria $\sqrt{\beta_1} = 1.0, 2.0, 3.0, 4.0$. I limiti di specifica sono $USL = 70$ e $LSL = 10$ e $C_p = 1$ per tutti e quattro i processi. La figura 5 dimostra la forma delle quattro distribuzioni.

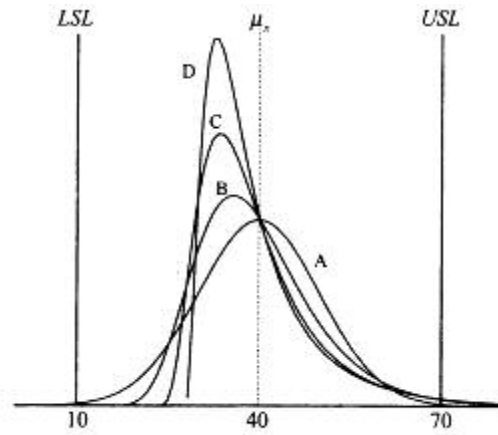


Figura 5

Si noti come la percentuale di prodotti non conformi aumenti quando l'asimmetria diminuisce rappresentando così C_p^{WSD} un fenomeno. Infatti questo diminuisce quando l'asimmetria aumenta anche se C_p rimane costante.

La stima di questo Indice è data da

$$\hat{C}_p^{WSD} = \frac{USL - LSL}{6 \cdot \hat{D}_x \cdot s_x}$$

con $\hat{D}_x = 1 + |1 - 2\hat{P}_x|$ e con $\hat{P}_x = \frac{1}{n} \sum_{i=1}^n I(\bar{x} - x_i)$ dove $I(x) = 1$ se $x \geq 0$ mentre $I(x) = 0$ se $x < 0$.

4.5.3.3. L'Indice C_{pk} basato sul metodo WSD.

Consideriamo il caso in cui esista solo un limite di specifica. Sotto il metodo WSD, gli Indici di Capacità inferiore e superiore sono definiti così:

$$C_{pku}^{WSD} = \frac{USL - \mu_x}{3 \cdot 2\sigma_u^W} = \frac{USL - \mu_x}{6P_x\sigma_x} \quad \text{e} \quad C_{pkl}^{WSD} = \frac{\mu_x - LSL}{3 \cdot 2\sigma_L^W} = \frac{\mu_x - LSL}{6(1 - P_x)\sigma_x}$$

Per il caso in cui i limiti di specifica siano due, l'Indice di Capacità C_{pk}^{WSD} è definito come:

$$C_{pk}^{WSD} = \min(C_{pku}^{WSD}, C_{pkl}^{WSD}) = \min\left\{ \frac{USL - \mu_x}{6 \cdot P_x \sigma_x}, \frac{\mu_x - LSL}{6 \cdot (1 - P_x) \sigma_x} \right\}$$

Da notare come, nel caso di distribuzioni simmetriche, C_{pk}^{WSD} si riduca a C_{pk} , e se $\mu_x = m = \frac{1}{2}(USL + LSL)$, C_{pk}^{WSD} si riduca a C_p^{WSD} .

Per fare un esempio, nella seguente tabella viene mostrato il C_{pk}^{WSD} e gli altri Indici per tre distribuzioni lognormali aventi la stessa forma e scala, ma differente posizione. Le medie dei processi A, B e C sono rispettivamente $\mu_x = 40, 35, 45$ ed hanno in comune $\sigma_x = 10, \sqrt{\beta_1} = 2.0$, la probabilità $P_x = 0.61$, $USL = 70$ e $LSL = 10$.

Processo	μ_x	PNC	C_{pk}^{WSD}	C_{pk}	C_p^{WSD}	C_p
A	40	16358	0.82	1.00	0.82	1.00
B	35	10400	0.96	0.83	0.82	1.00
C	45	26806	0.68	0.83	0.82	1.00

La figura 6 mostra come il C_{pk}^{WSD} per il processo B, che produce il valore più piccolo dei prodotti non conformi, sia più grande di quello per il processo A e C. Mentre il C_{pk} indica che il processo A è più Capace, C_p^{WSD} e C_p rimangono costanti. Questo significa che C_{pk}^{WSD} descrive la Capacità di un processo più accuratamente quando la distribuzione presa in considerazione è asimmetrica e il processo non è centrato, $\mu_x \neq m = \frac{1}{2} (USL + LSL)$.

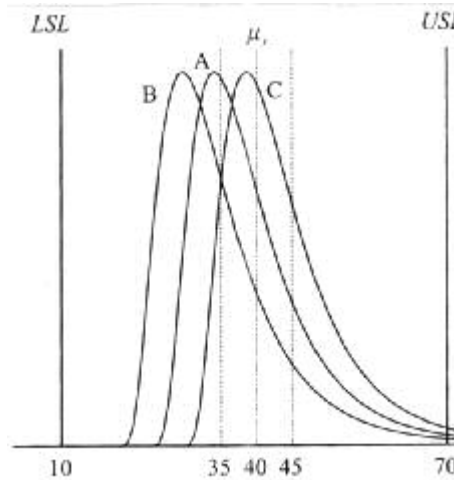


Figura 6

Per quanto riguarda la stima di C_{pk}^{WSD} , questa è data da

$$\hat{C}_{pk}^{WSD} = \min(\hat{C}_{pku}^{WSD}, \hat{C}_{pkl}^{WSD}) = \min\left\{ \frac{USL - \mu_x}{6 \cdot \hat{P}_x s_x}, \frac{\mu_x - LSL}{6 \cdot (1 - \hat{P}_x) s_x} \right\}$$

con $\hat{P}_x = \frac{1}{n} \sum_{i=1}^n I(\bar{x} - x_i)$ dove $I(x) = 1$ se $x \geq 0$ mentre $I(x) = 0$ se $x < 0$.

4.5.3.4. L'Indice C_{pm} e C_{pm}^* basato sul metodo WSD.

Il C_{pm} basato sul metodo WSD è definito come:

$$C_{pm}^{WSD} = \frac{USL - LSL}{6\sigma'_x} \min\left\{ \frac{1}{\sqrt{2P_T}}, \frac{1}{\sqrt{2(1-P_T)}} \right\}$$

dove $\sigma'_x = \sqrt{\sigma_x^2 + (\mu - T)^2}$.

Mentre, per quanto riguarda il C_{pm}^* basato sul WSD, la sua definizione è:

$$C_{pm}^{*WSD} = \min \left\{ \frac{USL - T}{3 \cdot \sigma'_x \sqrt{2P_T}}, \frac{LSL - T}{3 \cdot \sigma'_x \sqrt{2(1 - P_T)}} \right\}$$

Per quanto riguarda la stima di questi due Indici, questa sono date da:

$$\hat{C}_{pm}^{WSD} = \frac{USL - LSL}{6\hat{\sigma}'_x} \min \left\{ \frac{1}{\sqrt{2\hat{P}_T}}, \frac{1}{\sqrt{2 \cdot (1 - \hat{P}_T)}} \right\},$$

$$\hat{C}_{pm}^{*WSD} = \min \left\{ \frac{USL - T}{3 \cdot \hat{\sigma}'_x \sqrt{2\hat{P}_T}}, \frac{LSL - T}{3 \cdot \hat{\sigma}'_x \sqrt{2(1 - \hat{P}_T)}} \right\}$$

con $\hat{P}_T = \frac{1}{n} \sum_{i=1}^n I(T - X_i)$ e con $\hat{\sigma}'_x = \sqrt{s_x^2 + (\mu - T)^2}$.

4.5.3.5. Performance degli Indici di Capacità WSD.

Choi e Bai hanno quindi confrontato numericamente gli Indici di Capacità WSD con gli Indici standard di base e quelli ricavati dal metodo di Clements nel caso delle distribuzioni Normali, Weibull, Lognormale e Gamma. Queste distribuzioni sono state prese in considerazione perché rappresentano una varietà di forme, dalla simmetrica a quella molto asimmetrica.

Sono state quindi calcolate due tabelle che danno i valori degli Indici C_p^{WSD} e C_p^{*WSD} , C'_p e C'_{pk} di Clements, C_p e C_{pk} quando la numerosità del campione è n e quando l'asimmetria aumenta. Questi Indici sono stati ottenuti attraverso una simulazione di Monte Carlo quando n è finito e i parametri sono: $USL = 3$, $LSL = -3$, con le distribuzioni che hanno lo stesso valore di $\mu_x = 0$ e $\sigma_x = 1$ per la prima tabella mentre $\mu_x = 1$ e $\sigma_x = 1$ per la seconda tabella, cosicché $C_p = 1$ sempre per la prima tabella mentre $C_{pk} = 1$ sempre per la seconda tabella. Hanno infine calcolato anche percentuale NC.

Dopo aver ricavato queste tabelle, sono giunti alle seguenti conclusioni:

- a) Gli Indici di Capacità WSD sono migliori rispetto agli altri per le popolazioni asimmetriche.
- b) Quando i parametri del processo sono conosciuti, C_p^{WSD} e C'_p diminuiscono quando l'asimmetria aumenta anche se C_p^{*WSD} descrive meglio la Capacità del processo rispetto a C'_p e C_p . Nel caso in cui i tutti i parametri siano conosciuti, tutti gli Indici di Capacità sovrastimano la Capacità del processo, specialmente quando la numerosità campionaria è piccola, e questo è più sottolineato in C'_p e C_p .

Come si può notare quindi, C_p^{WSD} sottostima la vera Capacità del processo quando il valore di $|USL - LSL|$ è piccolo e al contrario la precisione è migliore quando $|USL - LSL|$ è circa 6σ .

- c) In generale C'_{pk} sottostima la Capacità del processo quando invece gli altri la sovrastimano, comunque C_{pk}^{WSD} è il migliore tra gli Indici per determinare la Capacità di un processo.
- d) La distorsione di \hat{C}_{pk}^{WSD} e di \hat{C}_p^{WSD} è piccola e non supera il 10% del varo valore di C_{pk}^{WSD} e di C_p^{WSD} , eccetto quando n è piccolo e $\sqrt{\beta_1}$ è elevato
- e) \hat{C}_p^{WSD} sovrastima il vero valore di C_p^{WSD} tranne quando la distribuzione è normale. Anche \hat{C}_{pk}^{WSD} sovrastima C_{pk}^{WSD} tranne quando $\sqrt{\beta_1}$ è piccola e $\mu_x = \frac{1}{2}(USL + LSL)$
- f) Per dati valori di $\sqrt{\beta_1}$ la relativa distorsione diminuisce con n , mentre per dati valori di n , la distorsione diminuisce quando $\sqrt{\beta_1}$ aumenta.

4.5.4. Gli Indici di Capacità basati sul metodo della varianza pesata (WV).

Il metodo visto precedentemente basato sulla deviazione standard di Bai e Choi è un po' differente dal metodo basato sulla varianza pesata, anche se comunque all'interno di entrambi è riposta la stessa idea, quella di dividere dalla sua media una distribuzione asimmetrica in due distribuzioni Normali.

Per una popolazione con media μ e deviazione standard σ , ci sono n_1 osservazioni su n che sono sotto la media e n_2 osservazioni su n che sono sopra la media. Le due nuove distribuzioni, trovate usando rispettivamente n_1 e n_2 , hanno la stessa media μ ma differente deviazione standard, σ_1 e σ_2 . Se la distribuzione è simmetrica, allora $n_1 = n_2 = n/2$ e $\sigma_1 = \sigma_2 = \sigma$. Dal momento che queste, insieme a μ , sono sconosciute, devono essere stimate rispettivamente con s_1 e s_2 , definite come:

$$s_1^2 = \frac{2 \sum_{i=1}^n (x_i - \bar{x})^2}{2n_1 - 1} \quad \text{e} \quad s_2^2 = \frac{2 \sum_{i=1}^n (x_i - \bar{x})^2}{2n_2 - 1}$$

D'altro canto devono essere trovate anche le deviazioni standard dal valore obiettivo, s_{T1} e s_{T2} , definite come:

$$s_{T1}^2 = \frac{2 \sum_{i=1}^n (x_i - T)^2}{2n_1} \quad \text{e} \quad s_{T2}^2 = \frac{2 \sum_{i=1}^n (x_i - T)^2}{2n_2}$$

$$= \frac{2n_1 - 1}{2n_1} \cdot s_1^2 + (\bar{x} - T)^2 \quad = \frac{2n_2 - 1}{2n_2} \cdot s_2^2 + (\bar{x} - T)^2$$

dove T è il valore obiettivo.

Quindi, dopo aver definito le stime delle deviazioni standard con il metodo della varianza pesata, queste possono essere collocate all'interno del denominatore dei quattro Indici fondamentali C_p , C_{pk} , C_{pm} e C_{pmk} , trovando quindi la loro stima.

Il denominatore originale di C_p è $\pm 3\sigma$, mentre in quello proveniente dal metodo della varianza pesata il è $3\sigma_1 + 3\sigma_2$. Quindi una stima di questo Indice con il metodo della varianza pesata è:

$$\hat{C}_p^{WV} = \frac{USL - LSL}{3(\sigma_1 + \sigma_2)}$$

Mentre, per quanto riguarda la stima di C_{pk} con il metodo della varianza pesata, ho che:

$$\hat{C}_{pk}^{WV} = \min\left\{\frac{USL - \bar{x}}{3 \cdot s_2}, \frac{\bar{x} - LSL}{3 \cdot s_1}\right\}$$

e per la stima di C_{pm} e C_{pmk} ho che:

$$C_{pm}^{WV} = \min\left\{\frac{USL - T}{3 \cdot s_{T2}}, \frac{T - LSL}{3 \cdot s_{T1}}\right\}$$

$$C_{pmk}^{WV} = \min\left\{\frac{USL - \bar{x}}{3 \cdot s_{T2}}, \frac{\bar{x} - LSL}{3 \cdot s_{T1}}\right\}$$

Il vantaggio, quindi, di usare gli Indici di Capacità basati sul metodo della varianza pesata piuttosto che quelli basati sulle distribuzioni normali è che questi riflettono l'asimmetria e la curtosi di una distribuzione. Infatti le loro deviazioni standard campionarie, s_1 e s_2 , e le deviazioni standard dal valore obiettivi, s_{T1} e s_{T2} , sono sistemate da differenti valore dell'asimmetria e della curtosi.

4.5.5. L'Indice generalizzato $C_{Np}(u, v)$.

4.5.5.1. Definizione di $C_{Np}(u, v)$.

Cominciamo con il dare la definizione dell'Indice $C_p(u, v)$. Questo Indice è stato costruito da Vannmam (1995) per dare una generalizzazione ai quattro Indici C_p , C_{pk} , C_{pm} e C_{pmk} , ed è stato definito come

$$C_p(u, v) = \frac{d - u|\mu - m|}{3\sqrt{\sigma^2 + v(u - T)^2}}$$

dove μ è la media, σ è la deviazione standard, $d = (USL - LSL)/2$, $m = (USL + LSL)/2$, T è il valore obiettivo con $u, v \geq 0$. È facile verificare che $C_p(0,0) = C_p$, $C_p(1,0) = C_{pk}$, $C_p(0,1) = C_{pm}$ e $C_p(1,1) = C_{pmk}$.

Una stima stabile e precisa di $C_p(u,v)$ può essere ottenuta sostituendo, come al solito, μ con \bar{x} e σ con s . Ma, per il caso di distribuzioni Non-Normali, queste stime diventano molto instabili dal momento che la distribuzione delle varianze campionarie, s^2 , è sensibile alla deviazione dalla normalità.

Per quanto riguarda la percentuale di prodotti non conformi NC ($p\%$), questa è minore o uguale a $200\Phi(-3 C_p(u,v))$, sottolineando come $C_p(u,v)$ sia limitato quando σ tende a zero. L'espressione di $C_p(u,v)$ dimostra anche che, se si vuole che l'Indice sia sensibile alla deviazione di μ da T , allora u e v devono essere elevati.

A questo punto, per trattare il caso di processi distribuiti Non-Normalmente, consideriamo la seguente generalizzazione $C_{Np}(u,v)$, definita come:

$$C_{Np}(u,v) = \frac{d - u|M - m|}{3\sqrt{\left[\frac{F_{99.865} - F_{0.135}}{6}\right]^2 + v(M - T)^2}}$$

dove F_α è l' α percentile, M è la mediana della distribuzione, $m = \frac{1}{2}(USL + LSL)$ è il punto centrale tra i limite inferiore e superiore e $u, v \geq 0$.

È quindi possibile notare come, rispetto alla formula di $C_{Np}(u,v)$, al posto di μ ci sia la mediana e al posto di σ ci sia $(F_{99.865} - F_{0.135})/6$, questo perché si vuole imitare quella proprietà della normale in cui $Prob(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = 0.0027$, ricavando così che, se $C_{Np}(u,v) = 1$ (assumendo che il processo sia centrato bene e nel valore obiettivo), la probabilità che il processo sia fuori dai limiti di specifica è necessariamente piccola. È possibile anche notare come la mediana M sia una misura della centralità più robusta rispetto alla media, particolarmente per distribuzioni asimmetriche e con code pesanti.

Prendendo quindi $(u,v) = (0,0)$, $(0,1)$, $(1,0)$ e $(1,1)$, si ottengono come visto le seguenti generalizzazioni dei quattro Indici fondamentali per le distribuzioni Non-Normali, C_{Np} , C_{Npk} , C_{Npm} e C_{Npmk} . La loro definizione è:

$$C_{Np} = \frac{USL - LSL}{F_{99.865} - F_{0.135}}$$

$$C_{Npk} = \min \left\{ \frac{USL - M}{\left[\frac{F_{99.865} - F_{0.135}}{2}\right]}, \frac{M - LSL}{\left[\frac{F_{99.865} - F_{0.135}}{2}\right]} \right\}$$

$$C_{Npm} = \frac{USL - LSL}{6\sqrt{\left[\frac{F_{99.865} - F_{0.135}}{6}\right]^2 + (M - T)^2}}$$

$$C_{Npmk} = \min \left\{ \frac{USL - M}{3 \sqrt{\left[\frac{F_{99.865} - F_{0.135}}{6} \right]^2 + (M - T)^2}}, \frac{M - LSL}{3 \sqrt{\left[\frac{F_{99.865} - F_{0.135}}{6} \right]^2 + (M - T)^2}} \right\}$$

Classificando questi quattro Indici, dal più sensibile al meno sensibile, in termini di sensibilità alla deviazione della media del processo dal valore obiettivo, si ottiene:

- (1) C_{Npmk} (2) C_{Npm} (3) C_{Npk} (4) C_{Np} .

Nel caso particolare e in cui la distribuzione sia normale, si ha che $M = \mu$ e $(F_{99.865} - F_{0.135}) = 6\sigma$, ottenendo quindi $C_{Np} = C_p$, $C_{Npk} = C_{pk}$, $C_{Npm} = C_{pm}$ e $C_{Npmk} = C_{pkm}$.

4.5.5.2. Stima di $C_{Np}(u,v)$.

Pearn e Chen (1995) hanno proposto uno stimatore di $C_{Np}(u,v)$ che si basa nell'utilizzo della distribuzione di Pearson, applicando quindi il metodo di Clements e sostituendo il 6σ con $U_p - L_p$ (vedere paragrafo 4.2.1). Lo stimatore è di conseguenza:

$$C_{Np}(u,v) = \frac{d - u|\hat{M} - m|}{3 \sqrt{\left[\frac{U_p - L_p}{6} \right]^2 + v(\hat{M} - T)^2}}$$

dove U_p e L_p stimano rispettivamente il percentile 99.865 e 0.135 della distribuzione.

Basato invece sempre sui percentili campionari, Chang e Lu (1994) hanno presentato un metodo differente rispetto al precedente per il calcolo di $F_{99.865}$, $F_{0.135}$ e della mediana M . Questo metodo è essenzialmente basato sui percentili campionari che vengono calcolati usando l'interpolazione, ottenendo le seguenti stime:

$$C_{Np}(u,v) = \frac{d - u|\hat{M} - m|}{3 \sqrt{\left[\frac{\hat{F}_{99.865} - \hat{F}_{0.135}}{6} \right]^2 + v(\hat{M} - T)^2}}$$

dove

$$\hat{F}_{99.865} = x_{(R_1)} + ([R_1] - R_1) \cdot (x_{(R_1+1)} - x_{R_1}) \quad \text{con} \quad R_1 = \frac{(99.865 \cdot n) + 0.135}{100}$$

$$\hat{F}_{0.135} = x_{(R_2)} + ([R_2] - R_2) \cdot (x_{(R_2+1)} - x_{R_2}) \quad \text{con} \quad R_2 = \frac{(0.135 \cdot n) + 99.865}{100}$$

$$\hat{M} = x_{R_3} + ([R_3] - R_3) \cdot (x_{(R_3+1)} - x_{R_3}) \quad \text{con} \quad R_3 = \frac{n+1}{2}$$

e dove $[R]$ è definito come il numero intero più grande minore o uguale a R , mentre x_i è definito come il termine i -esimo della serie di dati.

CAPITOLO 5

LE PRINCIPALI DISTRIBUZIONI NON - NORMALI

5.1. La distribuzione Gamma ($\Gamma(p, \lambda)$).

La distribuzione Gamma è una distribuzione che inizia dall'origine e si sviluppa secondo una curva flessibile, assume solo valori positivi, è asimmetrica a destra e la sua funzione di densità è:

$$f(x) = \frac{1}{\Gamma(p)} \cdot \lambda^p \cdot x^{p-1} \cdot e^{-\lambda x} \quad \text{con } 0 \leq x < \infty$$

dove $p > 0$ è il parametro di forma, $\lambda > 0$ è il parametro inversamente proporzionale alla media della distribuzione, $1/\lambda$ è detto parametro di scala, e tutti questi parametri non sono solo numeri interi ma anche razionali. Per quanto riguarda invece $\Gamma(p)$, questa è la funzione della Gamma completa ed è definita come

$$\Gamma(p) = \int_0^{\infty} t^{p-1} \cdot e^{-t} dt$$

Dalla funzione di densità, può essere ricavata la funzione di ripartizione:

$$F(x) = \text{Pr ob}[\Gamma(p, \lambda) \leq x] = \frac{\Gamma_x(p)}{\Gamma(p)}$$

con $\Gamma(p)$ che è la funzione Gamma completa mentre $\Gamma_x(p)$ è la funzione Gamma incompleta, definita come

$$\Gamma_x(p) = \int_0^x t^{p-1} \cdot e^{-t} dt$$

Per quanto riguarda il momento r-esimo attorno l'origine, questo è pari a

$$E(\Gamma(p, \lambda))^r = \frac{p^r \cdot \Gamma(\lambda + r)}{\Gamma(\lambda)}$$

ottenendo infine, per dati valori di r, i seguenti momenti:

$$E(\Gamma(p, \lambda)) = \frac{p}{\lambda}$$

$$\text{Var}(\Gamma(p, \lambda)) = \frac{p}{\lambda^2}$$

$$\sqrt{\beta_1} = \frac{2}{\sqrt{p}}$$

$$\beta_2 = 3 + \frac{6}{p}$$

$$\text{Moda} = \frac{p-1}{\lambda} \quad \text{con } p > 1$$

È possibile infine sottolineare le proprietà e le relazioni che la $\Gamma(p, \lambda)$ ha con le altre distribuzioni esistenti:

- a) La distribuzione Gamma standard è $\Gamma(1, \lambda)$
- b) La distribuzione Esponenziale è un caso particolare della Gamma, ovvero

$$\text{Exp}(\lambda) = \Gamma(1, \lambda)$$

- c) La distribuzione Chi – Quadro con n gradi di libertà è un caso particolare della Gamma, ovvero

$$\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

- d) Se $X_1 \approx \Gamma(p_1, \lambda)$ e $X_2 \approx \Gamma(p_2, \lambda)$ sono indipendenti, allora

$$x_1 + x_2 \approx \Gamma(p_1 + p_2, \lambda)$$

- e) Per quanto riguarda le stime dei parametri p e λ , dato un campione (X_1, \dots, X_n) e trovato \bar{X} ed s, queste possono essere ricavate con il metodo dei momenti, ovvero

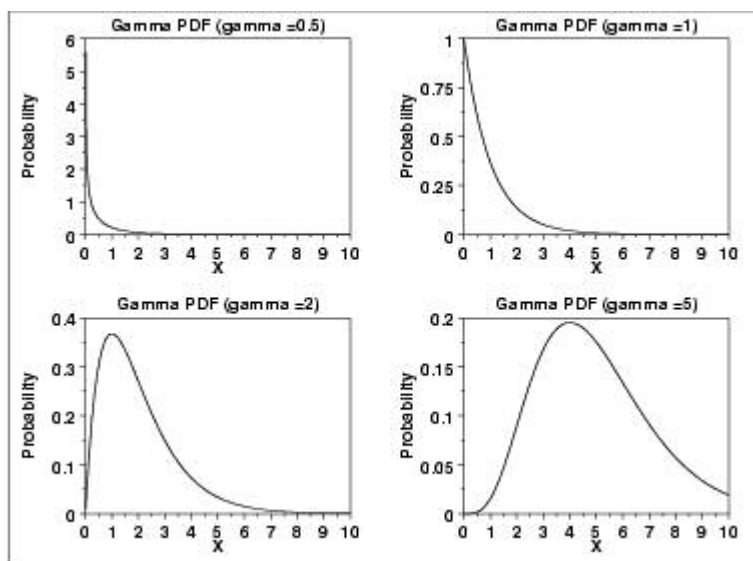
$$\bar{X} = \frac{p}{\lambda} \quad \text{e} \quad s^2 = \frac{p}{\lambda^2}$$

da cui

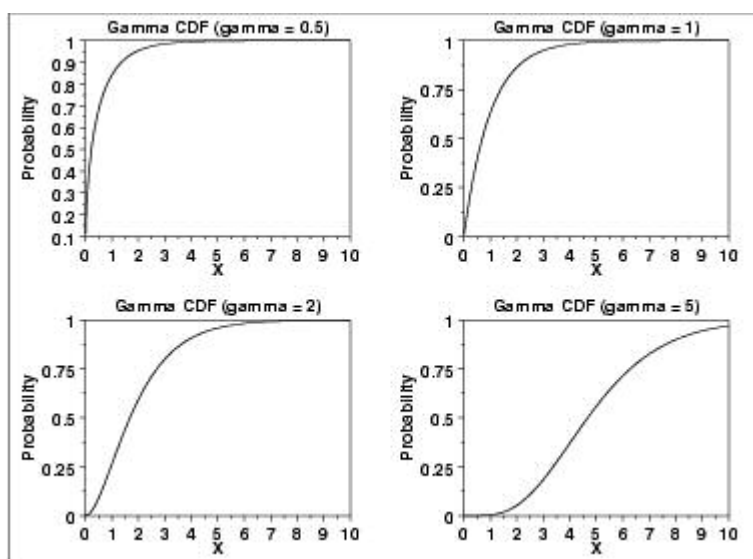
$$\hat{\lambda} = \frac{\bar{X}}{s^2} \quad \text{e} \quad \hat{p} = \frac{\bar{X}^2}{s^2}$$

- f) I percentili della distribuzione Gamma, come altre distribuzioni, non sono facili da ottenere e quindi, per facilitare i calcoli, sono stati tabulati. Grazie a queste tabulazioni, il calcolo dei percentili per la determinazione degli intervalli di confidenza e per la verifica d'ipotesi risulta facile.

g) Dei plot della funzione di densità per 4 determinati valori di p sono:



h) Dei plot della funzione di ripartizione per 4 determinati valori di p sono:



5.2. La distribuzione Chi – Quadro (χ^2_n).

Le più importanti applicazioni della distribuzione Chi – Quadro provengono dal fatto che questa è la distribuzione della somma dei quadrati di un numero di variabili Normali. Dove una serie di dati è rappresentata da un modello teorico, la distribuzione Chi – Quadro può essere usata per determinare la bontà dell’adattamento tra i dati osservati e i valori previsti dal modello, soggetto però a delle differenze essendo distribuito Normalmente.

Date infatti X_1, \dots, X_n variabili che si distribuiscono come delle $N(\mu_i, \sigma_i^2)$ allora

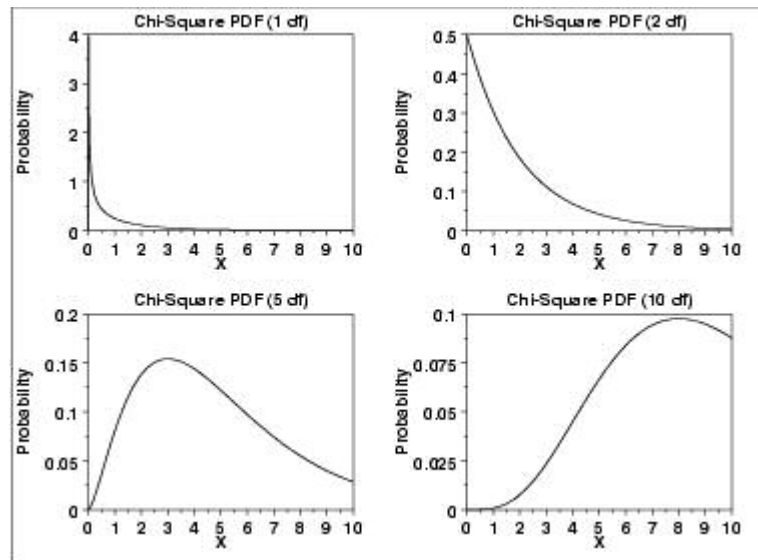
$$\chi_n^2 = \sum_{i=1}^n \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \approx \chi_n^2$$

dove n sono i gradi di libertà, che sono il numero di modi diversi in cui possono scegliersi n valori casuali da una popolazione.

Le distribuzione Chi – Quadro con n gradi di libertà è definita come χ_n^2 e la sua funzione di densità è:

$$f(x) = \frac{e^{-x/2} \cdot x^{(n-2)/2}}{2^{n/2} \cdot \Gamma(n/2)} \quad \text{con } 0 \leq x \leq \infty$$

dove $\Gamma(\cdot)$ è la funzione Gamma e dove la sua forma ed il suo centro dipendono dal numero di gradi di libertà della variabile. Dei plot della funzione di densità per 4 determinati valori di n sono:

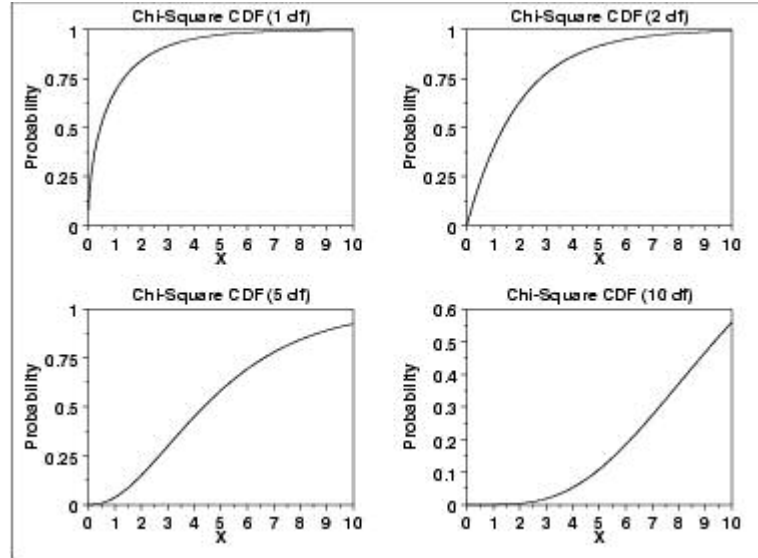


Dalla funzione di densità, per $n > 0$, può essere ricavata la funzione di ripartizione:

$$F(x) = \Pr ob[\chi_n^2 \leq x] = \frac{\Gamma_{x/2}(\frac{n}{2})}{\Gamma(\frac{n}{2})}$$

dove $\Gamma(p)$ è la funzione Gamma completa e $\Gamma_x(p)$ è la funzione Gamma incompleta.

Dei plot della funzione di ripartizione per 4 determinati valori di n sono:



Per quanto riguarda il momento r-esimo della Chi – Quadro attorno l'origine, questo è pari a

$$E(X_n^2)^r = \frac{2^r \Gamma(r + n/2)}{\Gamma(n/2)}$$

Dopo aver definito il momento r-esimo, per dati valori di r, possono essere ricavate i seguenti momenti:

$$E(X_n^2) = n$$

$$\text{Var}(X_n^2) = 2n$$

$$\sqrt{\beta_1(X_n^2)} = \sqrt{\frac{8}{n}}$$

$$\beta_2(X_n^2) = 3 + \frac{12}{n}$$

$$\text{Moda}(X_n^2) = n - 2 \text{ per } n \geq 2$$

$$\text{Mediana}(X_n^2) = n - 2/3$$

È possibile infine sottolineare le proprietà e le relazioni che la **Chi – Quadro** ha con le altre distribuzioni esistenti:

- a) Se non si conoscesse la media è quindi dovrebbe essere stimata, allora otterrei che

$$\sum_{i=1}^u \left(\frac{x_i - \bar{x}}{\sigma_i} \right)^2 \approx \chi_{n-1}^2$$

notando come ci siano infatti solo n - 1 valori x_i che possono scegliersi in modo completamente libero visto che, fissando la media campionaria, l'n-esimo valore risulta automaticamente determinato.

- b) Dal plot della funzione di densità della Chi – Quadro può essere notato come, quando n aumenta, la forma della curva di questa distribuzione diventi simile a

quella della Normale. Infatti, la distribuzione della Chi – Quadro dopo essere stata standardizzata, tende a quella della Normale, che è

$$\lim_{u \rightarrow \infty} \text{Pr ob} \left(\frac{\chi_n^2 - n}{\sqrt{2n}} \leq x \right) = \Phi(x)$$

dove $\Phi(\cdot)$ è la funzione cumulata della distribuzione Normale.

- c) La Chi – Quadro con n gradi di libertà appartiene alla famiglia delle distribuzioni Gamma con parametro di scala pari a 1/2 e gradi di libertà pari a n/2, o allo stesso modo è uguale al doppio di una Gamma con parametro di scala pari a 1 e gradi di libertà pari a n/2.

$$\chi_u^2 \cong \Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \cong 2 \cdot \Gamma(1, n/2)$$

- d) Due distribuzioni Chi – Quadro indipendenti con rispettivamente n e m gradi di libertà, sono legate alla distribuzione F con n e m gradi di libertà dalla seguente relazione:

$$F_{n,m} \cong \frac{\chi_n^2/n}{\chi_m^2/m}$$

- e) La distribuzione Chi – Quadro è legata alla t di Student con n gradi di libertà e alla $N(0,1)$ dalla seguente relazione:

$$t_n = \frac{N(0,1)}{\sqrt{\chi_n^2/n}}$$

- f) La somma di n distribuzioni Chi – Quadro è ancora una Chi –Quadro, infatti:

$$\sum_{i=1}^n \chi_{n_i}^2 \cong \chi_n^2 \text{ con } n = \sum_{i=1}^n n_i$$

- g) Se si conosce μ e quindi $X \approx N(\mu, \sigma^2)$, si può ricavare che

$$\sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 \approx \chi_n^2$$

mentre, se non si conosce μ e quindi $\bar{X} \approx N(\mu, \sigma^2/n)$, si può ricavare che

$$\sum_{i=1}^n \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \right)^2 \approx \chi_1^2$$

- h) Date n distribuzioni normali $N(\mu, \sigma^2)$ indipendenti, si ottiene che $\frac{(n-1) \cdot s^2}{\sigma^2} \approx \chi_{n-1}^2$.

- i) Consideriamo una serie di n_1 variabili $N(\mu_1, \sigma^2)$ indipendenti e una serie di n_2 variabili $N(\mu_2, \sigma^2)$ indipendenti, definiamo quindi le stime $\bar{x}_1, \bar{x}_2, s_1, s_2$, si ottiene che

$$(n_1 \cdot s_1^2 + n_2 \cdot s_2^2) \approx \chi_{n_1+n_2-2}^2$$

- j) La tabulazione della distribuzione Chi – Quadro dipende dai gradi di libertà e da α , infatti

$$\text{Pr ob}(\chi^2 \leq \chi_n^2) = 1 - \alpha$$

e quindi per trovare l'intervallo di confidenza per σ^2 all'interno del quale vi sia una proporzione $1 - \alpha$ di probabilità della distribuzione χ_{n-1}^2 , si ottiene

$$\text{Pr ob}(\chi_{1-\alpha/2}^2 < \chi_{n-1}^2 < \chi_{\alpha/2}^2) = 1 - \alpha$$

$$\text{Pr ob}(\chi_{1-\alpha/2}^2 < \frac{(n-1)}{\sigma^2} s^2 < \chi_{\alpha/2}^2) = 1 - \alpha$$

$$\text{Pr ob}\left(\frac{(n-1)}{\chi_{\alpha/2}^2} s^2 < \sigma^2 < \frac{(n-1)}{\chi_{1-\alpha/2}^2} s^2\right) = 1 - \alpha$$

5.3. La distribuzione t di Student (t_n).

La distribuzione t di Student è usata per testare se la differenza tra le medie di due campioni di osservazioni è statisticamente significativa. Un valore significativo di t porta a rifiutare l'ipotesi nulla, indicando quindi come le medie siano significativamente differenti.

La variabile casuale t è definita come il rapporto tra una variabile normale standardizzata e la radice quadrata di una variabile Chi – Quadro divisa per il numero dei suoi gradi di libertà.

Pertanto, date $Z \approx N(0,1)$ e $Y \approx \chi_n^2$, ottengo che la t di Student è definita come:

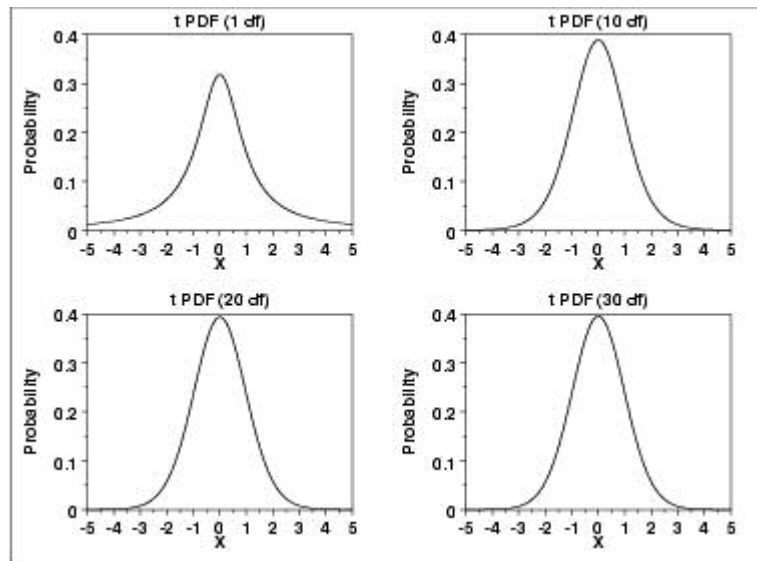
$$\frac{Z}{\sqrt{Y/n}} \approx t_n$$

con funzione di densità pari a:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \cdot \Gamma\left(\frac{n}{2}\right) \cdot \left(1 + \frac{x^2}{n}\right)^{(n+1)/2}} \quad \text{con } -\infty < x < +\infty$$

e con f(x) che presenta una forma simmetrica dipendente da n.

Dei plot della funzione di densità per 4 determinati valori di n sono:



La definizione della funzione di ripartizione è abbastanza complicata da determinare e definire, passando quindi alla definizione del momento r-esimo attorno l'origine, questo è

$$\begin{cases} E(t_n)^r = 0 & \text{per } r \text{ dispari} \\ E(t_n)^r = \frac{1 \cdot 3 \cdot 5 \dots (r-1) \cdot n^{r/2}}{(n-2) \cdot (n-4) \dots (n-r)} & \text{per } r \text{ pari e } n > r \end{cases}$$

ottenendo quindi per dati valori di n i seguenti momenti:

$$E(t_n) = 0$$

$$\text{Var}(t_n) = \frac{n}{n-2} \quad \text{per } n > 2$$

$$\sqrt{\beta_1} = 0 \quad \text{per } n > 3$$

$$\beta_2 = 3 + \frac{6}{n-4} \quad \text{per } n > 4$$

$$\text{Moda}(t_n) = 0$$

È possibile infine sottolineare le proprietà e le relazioni che la t di Student ha con le altre distribuzioni esistenti:

- a) Consideriamo delle distribuzioni $N(\mu, \sigma^2)$. Per standardizzare la media campionaria, usualmente viene usato s al posto di σ , ottenendo quindi che la distribuzione della media campionaria diventa una t di Student:

$$\frac{\bar{x} - \mu}{s/\sqrt{n-1}} \approx t_{n-1}$$

- b) Si osserva anche come, mentre la varianza campionaria di una t di Student dipende dai gradi di libertà n , quella di una Normale standard non dipende dalla numerosità campionaria n . La varianza di una t dipende quindi, a causa della variabilità presente in s (in particolare s risulta meno variabile per campioni grandi), dalla numerosità campionaria, osservando di conseguenza come per campioni con n piccolo la varianza sia più grande. Infatti è stato notato come la varianza tenda ad 1 quando n va all'infinito, potendo quindi approssimare quando n è elevato la $f(x)$ della t ad una $N(0,1)$. Questa approssimazione è ritenuta buona quando n è maggiore di 50 anche se la t di Student risulta, rispetto alla Normale standardizzata, più snella ma con le code più pesanti:

$$t_n \approx N(0,1) \quad \text{con } n > 50$$

- c) Si considerino una serie di n_1 variabili $N(\mu_1, \sigma^2)$ indipendenti e una serie di n_2 variabili $N(\mu_2, \sigma^2)$ indipendenti, definiamo quindi le stime \bar{x}_1 , \bar{x}_2 , s_1 , s_2 , si ottiene che

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 \cdot s_1^2 + n_2 \cdot s_2^2}{n_1 + n_2 - 2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \approx t_{n_1 + n_2 - 2}$$

- d) La tabulazione della distribuzione t di Student dipende dai gradi di libertà e da α , infatti

$$\text{Pr ob}(t \leq t_n) = 1 - \alpha$$

e quindi per trovare l'intervallo di confidenza per μ all'interno del quale vi sia una proporzione $1 - \alpha$ di probabilità della distribuzione t_{n-1} ,

si ottiene

$$\text{Prob}(-t_{\alpha/2} < t_{n-1} < t_{\alpha/2}) = 1 - \alpha$$

$$\text{Prob}\left(-t_{\alpha/2} < \frac{\bar{X} - \mu}{s/\sqrt{n}} < t_{\alpha/2}\right) = 1 - \alpha$$

$$\text{Prob}\left(\bar{X} - t_{\alpha/2} \frac{s}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

- e) Come visto nel punto b), quando n tende all'infinito, o più semplicemente è maggiore di 50, la t di Student tende a diventare una Normale standardizzata. Quindi l'intervallo

$$\bar{x} \pm t_{\alpha/2} \frac{s}{\sqrt{n}} \text{ tende a } \bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

osservando anche che $t_{0,025}$ si abbassa e tende a $z_{0,025} = 1.96$

- f) La distribuzione t è collegata a due distribuzioni $F_{n, n}$ indipendenti dalla relazione:

$$\frac{\sqrt{n}}{2} \cdot \left[\sqrt{F(1)_{n,n}} - \frac{1}{\sqrt{F(2)_{n,n}}} \right] \approx t_n$$

- g) Due distribuzioni Chi – Quadro con entrambe n gradi di libertà sono collegate alla distribuzione t dalla relazione:

$$\frac{\sqrt{n}}{2} \cdot \left[\frac{\chi_n^2(1) - \chi_n^2(2)}{\chi_n^2(1) \cdot \chi_n^2(2)} \right] \approx t_n$$

5.4. La distribuzione F di Fischer – Snedecor ($F_{m,n}$).

La distribuzione $F_{m,n}$ è il rapporto di due variabili indipendenti Chi – Quadro

$$F_{m,n} = \frac{\chi_m^2 / m}{\chi_n^2 / n}$$

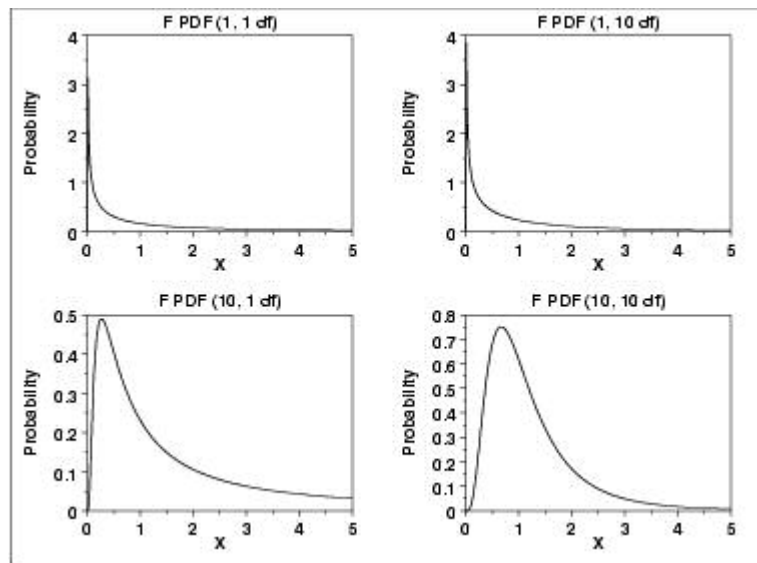
e, dal momento che la Chi – Quadro è la distribuzione della varianza tra i dati e il modello teorico, F determina una base per confrontare il rapporto delle varianze associate ai sottoinsiemi presi in considerazione con differenti elementi.

Si indicano quindi in $F_{m,n}$ per primi i gradi di libertà al numeratore e successivamente quelli che appaiono al denominatore. Infatti la F non è simmetrica in m e n.

La funzione di densità della distribuzione $F_{m,n}$ è:

$$f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2) \cdot \Gamma(n/2)} \cdot \left(\frac{m}{2}\right)^{m/2} \cdot x^{(m-2)/2} \cdot \left(1 + \frac{m}{n}x\right)^{-(m+n)/2} \quad \text{con } 0 \leq x < \infty$$

dove $\Gamma(p)$ è la funzione Gamma completa e dove dei plot della funzione di densità per 4 determinati valori di m ed n sono:



La funzione di ripartizione della F è

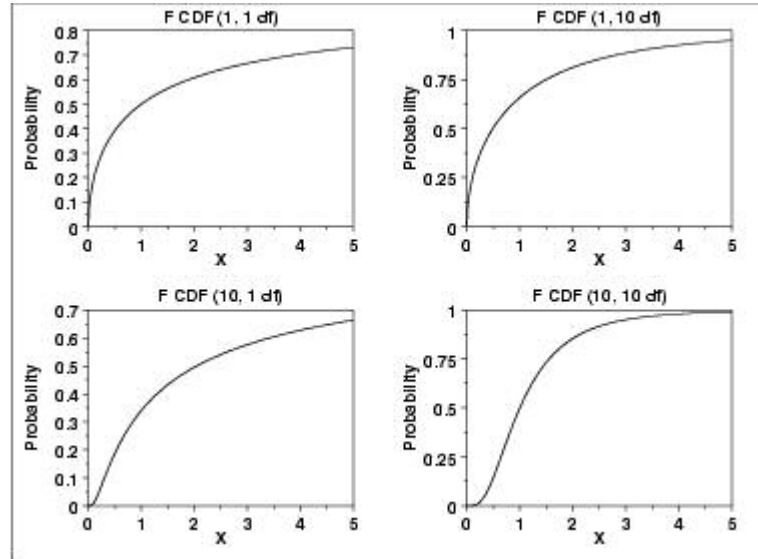
$$F(x) = \text{Pr ob}[F_{m,n} \leq x] = 1 - I_k\left(\frac{n}{2}, \frac{m}{2}\right)$$

dove $k = n/(n+m \cdot x)$ e I_k è la funzione della Beta incompleta, definita come

$$I_k(x, \alpha, \beta) = \frac{\int_0^x t^{\alpha-1} \cdot (1-t)^{\beta-1} dt}{B(\alpha, \beta)}$$

con $B(\alpha, \beta)$ che è la funzione della Beta completa.

Dei plot della funzione di ripartizione per 4 determinati valori di m ed n sono:



Per quanto riguarda il momento r-esimo della F attorno l'origine, questo è:

$$E(F_{m,n}^r) = \frac{(m/n)^r \cdot \Gamma(\frac{1}{2}n+r) \cdot \Gamma(\frac{1}{2}m+r)}{\Gamma(\frac{1}{2}m) \cdot \Gamma(\frac{1}{2}n)} \quad \text{per } m > 2r$$

e per dati valori di r si ottengono quindi i seguenti momenti:

$$E(F_{m,n}) = \frac{n}{n-2} \quad \text{per } n > 2$$

$$\text{Var}(F_{m,n}) = \frac{2n^2 \cdot (m+n-2)}{m(n-2)^2(n-4)} \quad \text{per } n > 4$$

$$\sqrt{\beta_1} = \frac{(2m+n-2) \cdot \sqrt{8(n-4)}}{\sqrt{m \cdot (n-6)} \cdot \sqrt{m+n-2}} \quad \text{per } n > 6$$

$$\beta_2 = 3 + \frac{12[(n-2)^2 \cdot (n-4) + m \cdot (m+n-2) \cdot (5n-22)]}{m \cdot (n-6) \cdot (n-8) \cdot (n+m-2)} \quad \text{per } n > 8$$

$$\text{Moda} = \frac{n \cdot (m-2)}{n \cdot (n+2)} \quad \text{per } m > 2$$

osservando come la media dipenda solamente dai gradi di libertà al denominatore mentre la varianza, l'asimmetria, la curtosi e la moda dipendono dai gradi di libertà al numeratore e al denominatore. Può pure essere osservato come quando i gradi di libertà m ed n aumentino, questa distribuzione tendi alla Normalità.

È possibile infine sottolineare le proprietà e le relazioni che la $F_{m,n}$ ha con le altre distribuzioni esistenti:

- a) Scambiano m ed n si ottiene la variabile $1/F_{m,n}$ che presenta n gradi di libertà al numeratore e che si distribuisce secondo una $F_{n,m}$. Infatti

$$\text{Pr ob}(F \leq F_{m,n}) = \text{Pr ob}\left(\frac{1}{F} \geq F_{n,m}\right)$$

da cui si ricava che il quantile $F_{n,m;1-\alpha} = \frac{1}{F_{m,n;\alpha}}$

- b) La distribuzione $F_{m, n}$ tende alla Chi – Quadro con m gradi di libertà quando n tende all'infinito:

$$F_{m,n} = \frac{1}{m} \cdot \chi_m^2 \quad \text{con } n \rightarrow \infty$$

- c) Il quantile $F_{1, n;\alpha}$ è uguale al quadrato del quantile della t di Student con n gradi di libertà e probabilità pari a $\frac{1}{2}(1+\alpha)$:

$$F_{1,n;\alpha} = (t_{n;1/2(1+\alpha)})^2$$

(ricordando comunque che i quantili sono tabulati all'interno di specifiche tavole)

- d) Si considerino due campioni $X \approx N(\mu_x, \sigma_x^2)$ e $Y \approx N(\mu_y, \sigma_y^2)$ di numerosità campionaria rispettivamente m ed n , ricordando che

$$\frac{(m-1) \cdot s_x^2}{\sigma_x^2} \approx \chi_{m-1}^2$$

$$\frac{(n-1) \cdot s_y^2}{\sigma_y^2} \approx \chi_{n-1}^2$$

si ricava che

$$F = \frac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2} \approx F_{m-1, n-1}$$

- e) Quindi, per quanto riguarda l'intervallo di confidenza, ho che:

$$\text{Pr ob}(F \leq F_{m-1, n-1; \alpha}) = 1 - \alpha$$

$$\text{Pr ob}(F_{m-1, n-1; 1-\alpha/2} \leq F \leq F_{m-1, n-1; \alpha/2}) = 1 - \alpha$$

$$\text{Pr ob} \left(F_{m-1, n-1; 1-\alpha/2} \leq \frac{s_x^2 / \sigma_x^2}{s_y^2 / \sigma_y^2} \leq F_{m-1, n-1; \alpha/2} \right) = 1 - \alpha$$

$$\text{Pr ob} \left(\frac{s_x^2}{(F_{m-1, n-1; \alpha/2}) \cdot s_y^2} \leq \frac{\sigma_x^2}{\sigma_y^2} \leq \frac{s_x^2}{(F_{m-1, n-1; 1-\alpha/2}) \cdot s_y^2} \right) = 1 - \alpha$$

osservando che se $\sigma_x^2 = \sigma_y^2$ allora

$$F = \frac{s_x^2}{s_y^2} \approx F_{m-1, n-1}$$

5.5. La distribuzione Lognormale (ln(x)).

La distribuzione Lognormale è applicabile alle variabili casuali che sono vincolate alla media (per il caso generale) e allo zero (per il caso standard) ma che hanno pochi valori molto grandi. L'applicazione di una trasformazione logaritmica per dati presi in considerazione può portare ad approssimare questi con una distribuzione Normale simmetrica, sebbene l'assenza di valori negativi potrebbe limitare la validità della trasformazione. Una distribuzione X è quindi Lognormalmente se $Y = \ln(X)$ è distribuita Normalmente, dove \ln è il logaritmo naturale.

La funzione di densità della Lognormale è

$$f(x) = \frac{\exp\left(-\frac{(\log(x - \mu/m))^2}{2\sigma^2}\right)}{(x - \mu) \cdot \sigma \cdot \sqrt{2\pi}} \quad \text{con } 0 \leq x \leq \infty; \quad x \geq \mu; \quad m > 0; \quad \sigma > 0$$

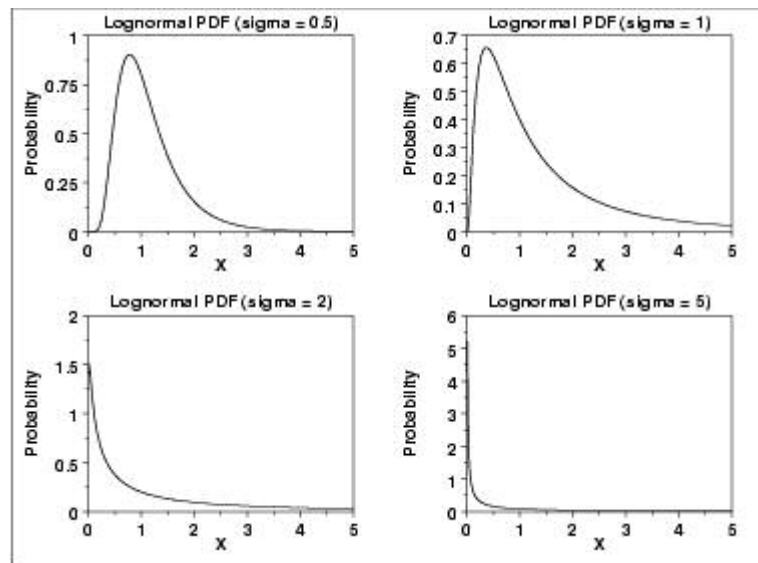
dove σ è il parametro di forma, μ è il parametro di locazione ed m è il parametro di scala. Esiste una relazione tra μ ed m , infatti queste due quantità sono legate dalla relazione

$$m = e^\mu \quad \text{e} \quad \mu = \ln(m)$$

Per il caso particolare $\mu = 0$ ed $m = 1$, si ottiene la distribuzione Lognormale standard che ha la funzione di densità pari a

$$f(x) = \frac{\exp\left(-\frac{(\log(x))^2}{2\sigma^2}\right)}{x \cdot \sigma \cdot \sqrt{2\pi}}$$

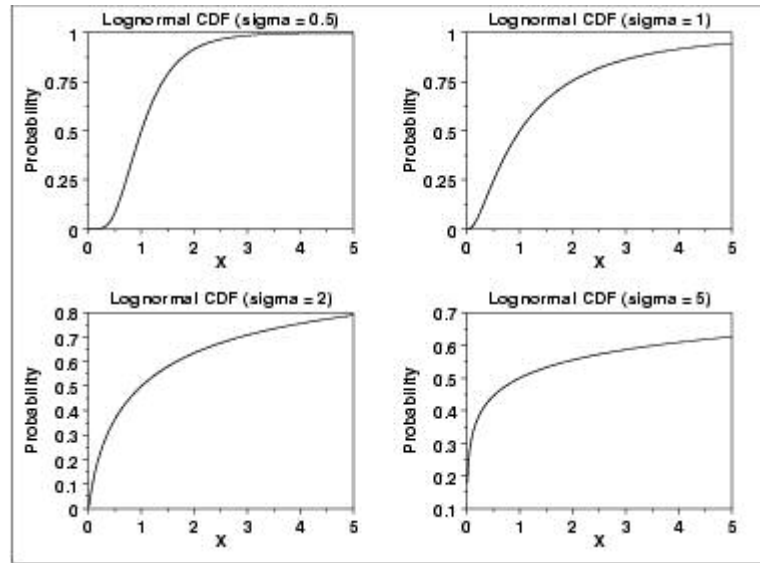
e dove dei plot di questa funzione di densità per 4 determinati valori di σ sono:



Per quanto riguarda la funzione di ripartizione della lognormale, questa è

$$F(x) = \Phi(\ln(x - \mu)/\sigma)$$

dove Φ la funzione di ripartizione della Normale standard ed è tabulata. Dei plot della funzione di ripartizione per 4 determinati valori di σ sono:



Per quanto riguarda il momento r-esimo attorno l'origine, la sua definizione generale è

$$E(\ln(m, \sigma))^r = m^r \cdot \exp\left(\frac{1}{2}r^2 \cdot \sigma^2\right) = \exp\left(r \cdot \mu + \frac{1}{2}r^2 \sigma^2\right)$$

e per dati valori di r si ottengono quindi i seguenti momenti:

$$E(\ln(X)) = m \cdot \exp\left(\frac{1}{2}\sigma^2\right)$$

$$\text{Var}(\ln(X)) = m^2 \cdot \exp(\sigma^2) \cdot (\exp(\sigma^2) - 1)$$

$$\sqrt{\beta_1} = (\exp(\sigma^2) + 2) \cdot \sqrt{\exp(\sigma^2) - 1}$$

$$\beta_2 = (\exp(\sigma^2))^4 + 2(\exp(\sigma^2))^3 + 3(\exp(\sigma^2))^2 - 3$$

$$\text{Moda} = m / (\exp(\sigma^2))$$

$$\text{Mediana} = m$$

È possibile infine sottolineare le proprietà e le relazioni che la $\ln(X)$ ha con le altre distribuzioni esistenti:

- a) La distribuzione Lognormale con parametri m e σ (quindi $\mu = 0$) è collegata alla distribuzione $N(\mu, \sigma)$ dalla seguente relazione:

$$\ln(m, \sigma) \approx \exp(N(\mu, \sigma)) \approx \exp(\mu + \sigma \cdot N(0,1)) \approx m \cdot \exp(\sigma \cdot N(0,1))$$

ora quindi

$$\log(\ln(m, \sigma)) \approx N(\mu, \sigma) \approx \mu + \sigma \cdot N(0,1)$$

- b) Date due distribuzioni Lognormali indipendenti, $\ln(\mu_1, \sigma_1)$ e $\ln(\mu_2, \sigma_2)$,

$$\ln(\mu_1, \sigma_1) \cdot \ln(\mu_2, \sigma_2) \approx \ln\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

$$\ln(\mu_1, \sigma_1) / \ln(\mu_2, \sigma_2) \approx \ln\left(\mu_1 - \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$

- c) Per quanto riguarda la stima dei tre parametri m , μ e σ ho che:

$$\hat{m} = \exp(\hat{\mu})$$

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n \log(x_i)$$

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (\log(x_i) - \hat{\mu})^2$$

se il parametro di scala μ è conosciuto è non deve essere stimato, allora questo può essere sottratto dalla serie originale dei dati prima di calcolare la stima dei parametri m e σ .

- d) I percentili della Lognormale posso essere ricavati dalla relazione

$$x_{\alpha} = \exp(\sigma \cdot \Phi^{-1}(\alpha))$$

dove Φ^{-1} è la funzione di ripartizione inversa della normale ed è tabulata.

5.6. La distribuzione Esponenziale ($\exp(\lambda)$).

La distribuzione Esponenziale è un caso particolare della Gamma, infatti queste due distribuzioni sono legate dalla relazione

$$\text{Exp}(\lambda) = \Gamma(1, \lambda)$$

e potrebbe rappresentare la distribuzione del tempo, in particolare quando la probabilità dell'evento che sta per accadere in un piccolo intervallo di tempo non varia attraverso il tempo, o anche potrebbe rappresentare la distribuzione del tempo tra degli eventi quando il numero di eventi in un qualche intervallo di tempo hanno una distribuzione di Poisson.

La funzione di densità della Esponenziale è:

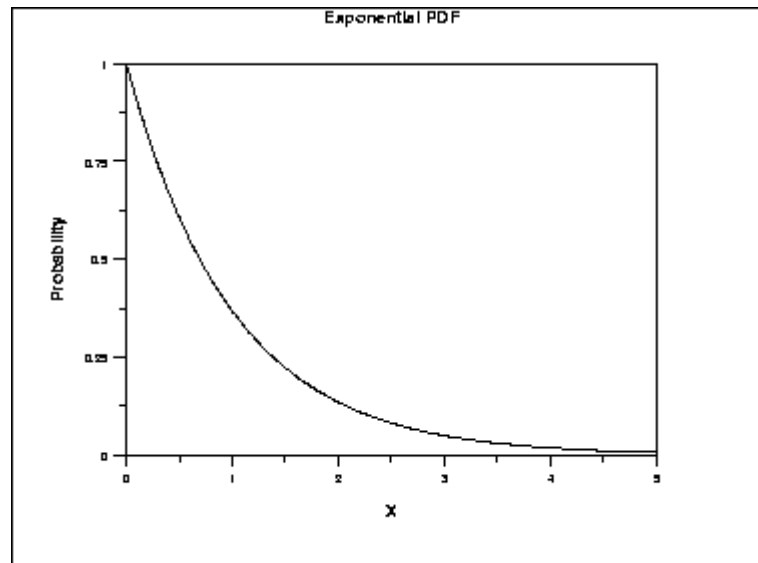
$$f(x) = \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right) \quad \text{con } 0 \leq x < \infty; \quad x > \mu; \quad \lambda > 0$$

dove μ è il parametro di locazione e σ è il parametro di scala. Spesso però il parametro di scala è espresso come $\lambda = 1/\sigma$ e solitamente viene studiato il caso particolare in cui $\mu = 0$, ottenendo così la distribuzione Esponenziale ad un solo parametro λ .

Nel caso in cui, oltre ad $\mu = 0$, $\sigma = 1$, si ottiene la distribuzione Esponenziale standard con funzione di densità pari a

$$f(x) = \exp(-x) \quad \text{con } x > 0$$

ed un plot di questa è:

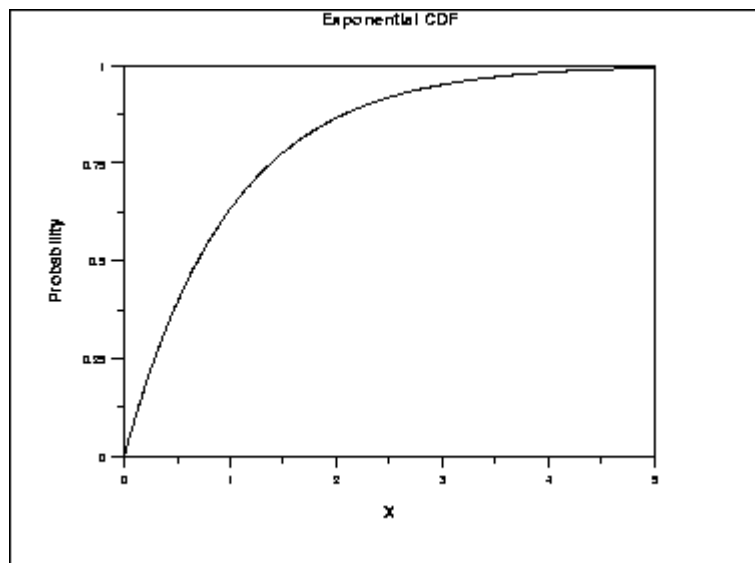


La distribuzione Esponenziale è l'unica distribuzione continua caratterizzata da un "salto di memoria".

La funzione di ripartizione è

$$F(x) = 1 - \exp(-\lambda x) \quad \text{con } x \geq 0; \quad \lambda > 0$$

e per il caso standard una rappresentazione grafica è:



Per quanto riguarda il momento r-esimo attorno l'origine, questo è pari a

$$E(\exp(\lambda))^r = r! \cdot (1/\lambda)^r$$

e per dati valori di r si ottengono quindi i seguenti momenti:

$$E(\exp(\lambda)) = 1/\lambda$$

$$\text{Var}(\exp(\lambda)) = (1/\lambda)^2$$

$$\sqrt{\beta_1} = 2$$

$$\beta_2 = 9$$

$$\text{Moda} = 0$$

$$\text{mediana} = 1/\lambda \cdot \log(2)$$

È possibile infine sottolineare le proprietà e le relazioni che la **Exp(λ)** ha con le altre distribuzioni esistenti:

- a) La distribuzione Esponenziale $\exp(\lambda)$ è un caso speciale della distribuzione di Weibull $W(b, c)$ con $c = 1$, parametro di scala:

$$\exp(\lambda) = W(\lambda, 1)$$

- b) La distribuzione Esponenziale $\exp(\lambda)$ è legata alla distribuzione rettangolare R dalla relazione:

$$\exp(\lambda) = -\lambda \cdot \log R$$

- c) La somma di c variabili Esponenziali indipendenti $\exp(\lambda)$ è una Gamma (p, λ) con λ parametro intero:

$$\sum_{i=1}^n (\exp(\lambda))_i = \Gamma(p, \lambda)$$

- d) La differenza di due variabili Esponenziali, rispettivamente $\exp_1(\lambda)$ ed $\exp_2(\lambda)$, è una distribuzione di Laplace con parametri $(0, \lambda)$, $L(0, \lambda)$:

$$L(0, \lambda) = \exp_1(\lambda) - \exp_2(\lambda)$$

- e) Per quanto riguarda la stima del parametro λ , questa è:

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

- f) I percentili della Esponenziale si ricavano dalla relazione

$$x_\alpha = -1/\lambda \cdot \ln(1 - \alpha)$$

CAPITOLO 6

CONFRONTO TRA GLI INDICI DI CAPACITA' TRAMITE SIMULAZIONE

6.1. Metodo di ricerca.

Kotz e Johnson hanno sottolineato come il problema fondamentale della Non - Normalità conduca verso la ricerca delle proprietà fondamentali degli Indici di Capacità e quindi dei loro stimatori attraverso l'uso di una varietà di distribuzione Non-Normali. Per avere quindi una base sicura, il metodo più corretto è quello di studiare un ampio range di distribuzioni Non-Normali, sia asimmetriche che simmetriche, per poi poter così confrontare i risultati ottenuti. Dal momento che il numero di alternative è infinito, solo un campione di questa infinità è possibile. Per quanto riguarda quindi le distribuzioni asimmetriche studierò la Gamma e la Chi-Quadro mentre per la distribuzione simmetrica la t di Student. Questo studio si basa sulla simulazione di 1000 campioni, variando la numerosità campionaria n (e precisamente $n = 50, 100, 150$ e 200) e i coefficienti di asimmetria e curtosi.

Prima di cominciare l'analisi di queste distribuzioni definisco la media del processo e stabilisco che dovrà essere pari a 3 per ogni distribuzione. Dopo aver definito la media del processo, cerco di definire i limiti specifica USL e LSL e questi saranno ricavati in modo che, sotto l'assunzione di Normalità e di $\mu = T$, il $C_{pk} = C_{pm} = C_{pmk} = 1$, risultando così il processo sotto controllo. Quindi considerando una Normale di $\mu = 3$ e di $\sigma = 1$, i limiti di specifica risulteranno pari a $LSL = 0$ e $USL = 6$ e il valore obiettivo $T = 3$.

Per la valutazione finale della bontà del processo viene usata la percentuale di prodotti non conformi (NC), percentuale che viene trasformata in un valore equivalente basato sulla distribuzione Normale, per l'appunto il C_{pk} equivalente.

6.2. Analisi della distribuzione Gamma(p, λ).

Definita quindi la media del processo e i limiti di specifica, cercherò rispettivamente 3 parametri p e λ in modo che la media della Gamma risulti pari a 3 ed i coefficienti di asimmetria e curtosi varino in ordine decrescente, ottenendo così:

Distribuzione	Media	Varianza	Asimmetria (√β ₁)	Curtosi (β ₂)	NC	C _{pk}
Γ (0.445,0.669)	3	1	3	16.48	0.38%	0.8900
Γ (1, 1)	3	1	2	9	0.25%	0.9366
Γ (4, 2)	3	1	1	4.5	0.23%	0.9466

Per ognuno dei tre casi da studiare simulo quindi 1000 campioni di numerosità campionaria 50, 100, 150 e 200 e per ognuno dei 1000 campioni calcolo i vari Indici di Capacità con il metodo di Clements, con il C_{jkp}, con il metodo delle varianze pesate, con il metodo delle deviazioni standard pesate, con il metodo di Wright e infine con il metodo generalizzato.

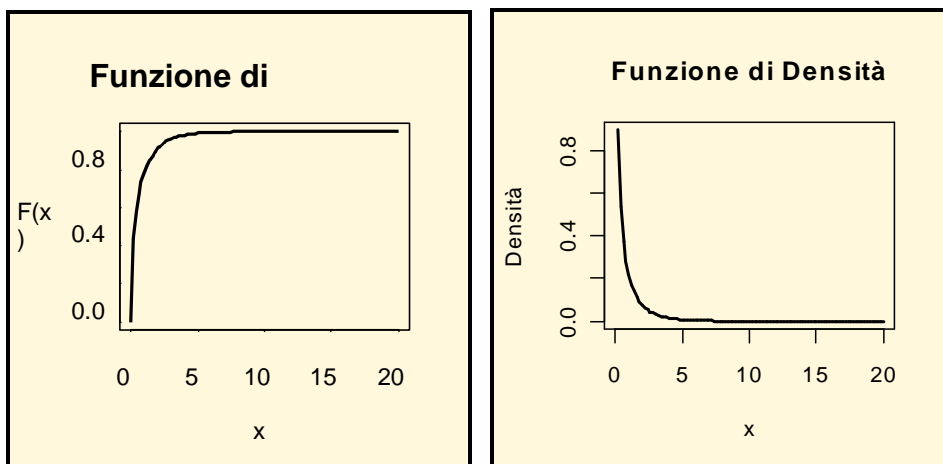
All'interno di ogni tabella viene rappresentato, per ogni numerosità campionaria, il valore dell'Indice e tra parentesi la sua deviazione standard, mentre la riga finale di ogni tabella rappresenta la differenza media percentuale con il C_{pk} equivalente. Da notare come il valore dell'Indice calcolato sia la media di tutti gli Indici calcolati per i mille campioni e come lo stesso valga anche per la deviazione standard.

Per quanto riguarda il calcolo del C_{pk} equivalente, questo è stato fatto secondo la definizione di Wierda (1994), che sviluppò una approssimazione del C_{pk} in funzione del rendimento (θ):

$$\tilde{C}_{pk} \cong \frac{1}{3} \Phi^{-1}(\theta)$$

Si può notare anche da questa tabella come al diminuire del coefficiente di asimmetria e curtosi diminuisca verso lo 0 anche la percentuale di prodotti non conformi, diventando quindi il nostro processo sempre più Capace e il C_{pk} equivalente tendente all' 1.

a) p = 0.445 , λ = 0.667



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.4790 (0.5161)	1.1908 (0.3422)	1.0605 (0.2869)	1.0000 (0.2511)
C_{pk}	0.9034 (0.3412)	0.7152 (0.2184)	0.6326 (0.1861)	0.5943 (0.1570)
C_{pm}	1.2353 (0.3175)	1.0590 (0.2051)	0.9770 (0.2098)	0.9166 (0.2444)
C_{pmk}	0.8413 (0.2780)	0.6876 (0.1965)	0.6206 (0.1579)	0.5754 (0.1415)
	24.59%	21.19%	19.74%	18.63%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.8949 (0.1800)	0.9061 (0.1372)	0.9123 (0.1120)	0.9222 (0.0993)
	0.49%	1.61%	2.23%	3.22%

3. *Il C_s di Wright:*

	n = 50	n = 100	n = 150	n = 200
C_s	0.1113 (0.0444)	0.0693 (0.0206)	0.0536 (0.0138)	0.0455 (0.0102)
	77.86%	82.06%	83.63%	84.44%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.8230 (0.2784)	0.7752 (0.1821)	0.7598 (0.1433)	0.7518 (0.1247)
C_{pk}^{WSD}	0.8320 (0.3118)	0.7808 (0.2068)	0.7642 (0.1619)	0.7544 (0.1413)
C_{pm}^{WSD}	0.7715 (0.1898)	0.7428 (0.1361)	0.7326 (0.1116)	0.7285 (0.0968)
	8.11%	12.37%	13.77%	14.50%

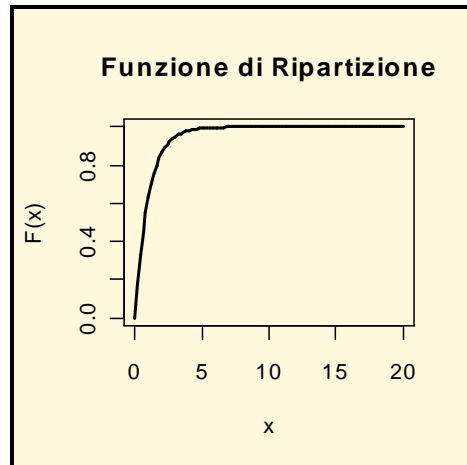
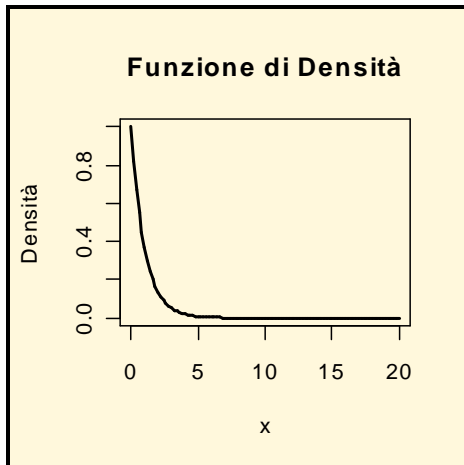
5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	1.0329 (0.2929)	0.9830 (0.1998)	0.9679 (0.1595)	0.9628 (0.1970)
C_{pk}^{WV}	0.7127 (0.2544)	0.6639 (0.1702)	0.6429 (0.1352)	0.6437 (0.1157)
C_{pm}^{WV}	0.7119 (0.2215)	0.6636 (0.1522)	0.6483 (0.1385)	0.6438 (0.1038)
C_{pmk}^{WV}	0.7183 (0.2473)	0.6673 (0.1684)	0.6513 (0.1347)	0.6455 (0.1152)
	16.74%	19.30%	19.98%	20.24%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.7007 (1.7448)	1.7104 (31.24)	0.4381 (0.7167)	0.3658 (0.2680)
C_{Npk}	0.3031 (0.7640)	0.7414 (13.47)	0.1896 (0.3099)	0.1587 (0.1172)
C_{Npm}	0.5959 (0.5171)	0.4713 (0.3920)	0.4131 (0.2952)	0.3609 (0.1963)
C_{Npmk}	0.2573 (0.3376)	0.2036 (0.3376)	0.1788 (0.3376)	0.1566 (0.3376)
	23.87%	51.86%	58.54%	62.94%

b) $p = 1, \lambda = 1$



1. Il $C_p, C_{pk}, C_{pm}, C_{pmk}$ con il metodo di Clements:

	n = 50	n = 100	n = 150	n = 200
C_p	1.4243 (0.3878)	1.2342 (0.2760)	1.1508 (0.2349)	1.0585 (0.2179)
C_{pk}	0.9488 (0.3160)	0.7986 (0.2093)	0.7354 (0.1738)	0.6679 (0.1571)
C_{pm}	1.2963 (0.2811)	1.1360 (0.2080)	1.0542 (0.1951)	1.0079 (0.1795)
C_{pmk}	0.9235 (0.2621)	0.7719 (0.1924)	0.6993 (0.1610)	0.6606 (0.1432)
	22.47%	19.98%	19.08%	18.51%

2. Il C_{jkr} :

	n = 50	n = 100	n = 150	n = 200
C_{jkr}	0.8413 (0.1179)	0.8553 (0.0735)	0.8520 (0.0578)	0.8514 (0.0544)
	9.52%	8.12%	8.45%	8.51%

3. Il C_s di Wright:

	n = 50	n = 100	n = 150	n = 200
C_s	0.1239 (0.0451)	0.0816 (0.0210)	0.0644 (0.0136)	0.0536 (0.0104)
	81.26%%	85.49%	87.21%	88.29%

4. Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.8452 (0.2227)	0.8350 (0.1475)	0.8203 (0.1218)	0.8060 (0.1076)
C_{pk}^{WSD}	0.8477 (0.2478)	0.8408 (0.1699)	0.8239 (0.1413)	0.8076 (0.1249)
C_{pm}^{WSD}	0.7309 (0.1348)	0.7311 (0.0941)	0.7239 (0.0790)	0.7149 (0.0697)
	12.86%	13.42%	14.71%	16.01%

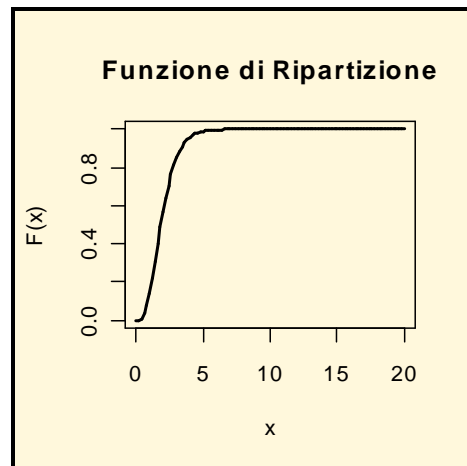
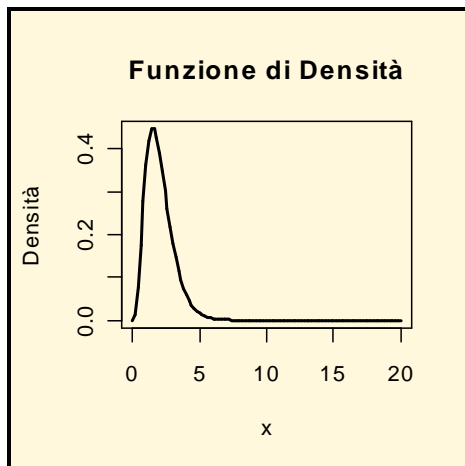
5. Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	1.0111 (0.2057)	1.0075 (0.1794)	0.9963 (0.1163)	0.9836 (0.1020)
C_{pk}^{WV}	0.7604 (0.2160)	0.7492 (0.1685)	0.7369 (0.1187)	0.7234 (0.1039)
C_{pm}^{WV}	0.7614 (0.1876)	0.7479 (0.1215)	0.7366 (0.1059)	0.7241 (0.0942)
C_{pmk}^{WV}	0.7654 (0.2114)	0.7521 (0.1376)	0.7389 (0.1180)	0.7248 (0.1033)
	14.93%	15.77%	16.42%	17.11%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.5557 (0.4041)	0.4629 (0.2654)	0.3806 (0.1069)	0.3721 (0.2185)
C_{Npk}	0.2460 (0.1768)	0.2054 (0.1568)	0.1702 (0.0480)	0.1667 (0.0983)
C_{Npm}	0.5437 (0.3357)	0.4363 (0.2589)	0.3796 (0.1051)	0.3694 (0.1638)
C_{Npmk}	0.2409 (0.3376)	0.1942 (0.3376)	0.1698 (0.3376)	0.1655 (0.3376)
	53.99%	61.18%	65.21%	66.81%

b) $p = 4, \lambda = 2$



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.3835 (0.2611)	1.2105 (0.1990)	1.1426 (0.1713)	1.1046 (0.1527)
C_{pk}	1.0933 (0.2984)	0.9195 (0.2128)	0.8531 (0.1765)	0.8182 (0.1540)
C_{pm}	1.3209 (0.2278)	1.1840 (0.1802)	1.1283 (0.1633)	1.0772 (0.1500)
C_{pmk}	1.0627 (0.2654)	0.9138 (0.1989)	0.8571 (0.1744)	0.8063 (0.1559)
	27.26%	14.04%	13.96%	13.89%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.7743 (0.0829)	0.7681 (0.0573)	0.7687 (0.0453)	0.7667 (0.0379)
	17.92%	17.84%	17.78%	17.98%

3. *Il C_s di Wright:*

	n = 50	n = 100	n = 150	n = 200
C_s	0.1769 (0.0948)	0.1101 (0.0321)	0.0880 (0.0195)	0.0748 (0.0136)
	76.66%	83.64%	85.85%	87.17%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.9045 (0.1672)	0.8884 (0.1241)	0.8971 (0.1375)	0.8936 (0.0900)
C_{pk}^{WSD}	0.8845 (0.1740)	0.8799 (0.1363)	0.8950 (0.1169)	0.8930 (0.1029)
C_{pm}^{WSD}	0.7154 (0.0938)	0.7075 (0.0663)	0.7114 (0.0541)	0.7098 (0.0460)
	11.17%	12.12%	11.20%	11.44%

5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	1.0123 (0.1398)	0.9976 (0.0968)	1.0017 (0.0789)	0.9988 (0.0672)
C_{pk}^{WV}	0.8636 (0.1806)	0.8412 (0.1245)	0.8447 (0.1026)	0.8403 (0.0875)
C_{pm}^{WV}	0.8648 (0.1596)	0.8441 (0.1107)	0.8444 (0.0910)	0.8399 (0.0770)
C_{pmk}^{WV}	0.8664 (0.1773)	0.8428 (0.1237)	0.8458 (0.1020)	0.8412 (0.0871)
	7.76%	9.07%	8.99%	9.26%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.5055 (0.3425)	0.4189 (0.1027)	0.3962 (0.0942)	0.3686 (0.0507)
C_{Npk}	0.2345 (0.1495)	0.1971 (0.0483)	0.1861 (0.0374)	0.1733 (0.0235)
C_{Npm}	0.4976 (0.2442)	0.4138 (0.1016)	0.3958 (0.0788)	0.3683 (0.0505)
C_{Npmk}	0.2311 (0.3376)	0.1971 (0.3376)	0.1860 (0.3376)	0.1732 (0.3376)
	57.94%	63.87%	65.55%	67.56%

6.3. Analisi della distribuzione Chi-Quadro (χ^2).

Definita anche per la Chi - Quadro la media del processo e i limiti di specifica, cercherò rispettivamente 3 parametri v in modo che la media della Chi-Quadro risulti pari a 3 ed i coefficienti di asimmetria e curtosi varino in ordine decrescente, ottenendo così:

Distribuzione	Media	Varianza	Asimmetri $a (\sqrt{\beta_1})$	Curtosi (β_2)	NC	C_{pk}
Chi-Quadro (0.4)	3	1	4.47	33	0.38%	0.8933
Chi-Quadro (0.5)	3	1	4	27	0.51%	0.8566
Chi-Quadro (0.6)	3	1	3.65	23	0.65%	0.8266

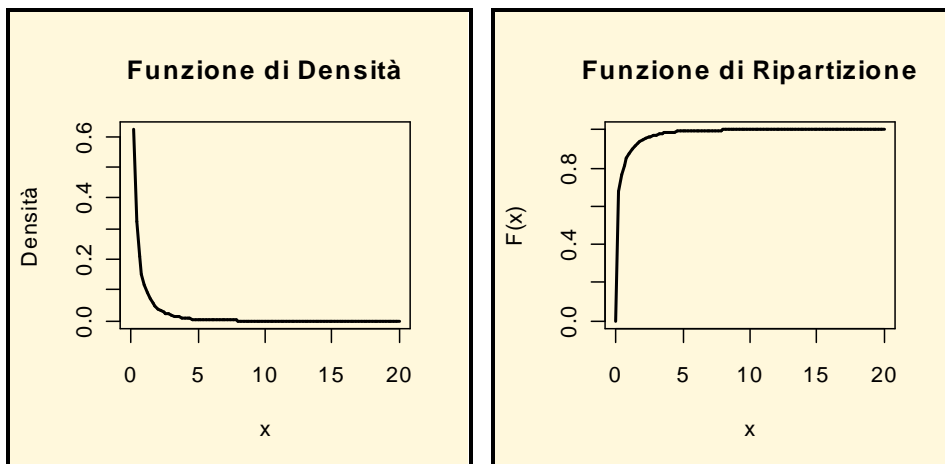
Per ognuno dei tre casi da studiare simulo quindi 1000 campioni di numerosità campionaria 50, 100, 150 e 200 e per ognuno dei 1000 campioni calcolo i vari Indici di Capacità con il metodo di Clements, con il $C_{j_{kp}}$, con il metodo delle varianze pesate, con il metodo delle deviazioni standard pesate, con il metodo di Wright e infine con il metodo generalizzato.

All'interno di ogni tabella viene rappresentato, per ogni numerosità campionaria, il valore dell'Indice e tra parentesi la sua deviazione standard. Da notare come il valore dell'Indice calcolato sia la media di tutti gli Indici calcolati per i mille campioni e come lo stesso valga anche per la deviazione standard.

Da notare come per la Chi - Quadro, dal momento che la varianza è pari a $2v$, per ottenere tre distribuzioni tutte di varianza pari a 1 ho preso come parametri v tre parametri tutti vicini tra di loro, precisamente intorno a 0.5. Quindi la varianza non sarà mai precisamente pari a 1 tranne per la Chi - Quadro con $v = 0.5$, ma vicino a 1 mentre i coefficienti di asimmetria e curtosi varieranno in ordine decrescente.

Altra cosa importante da sottolineare è come al diminuire dei coefficienti di asimmetria e curtosi, diversamente dalla Gamma, la percentuale di prodotti non conformi aumenti, diminuendo così il C_{pk} equivalente.

a) Chi - Quadro ($\chi^2_{0,4}$)



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.6886 (0.8373)	1.2985 (0.4898)	1.0976 (0.3635)	1.0112 (0.3217)
C_{pk}	0.9581 (0.4839)	0.7340 (0.2802)	0.6193 (0.2070)	0.5702 (0.1829)
C_{pm}	1.3777 (0.4251)	1.1062 (0.3190)	1.0126 (0.2704)	0.9340 (0.2551)
C_{pmk}	0.8994 (0.3641)	0.6785 (0.2319)	0.6102 (0.1878)	0.5565 (0.1703)
	34.44%	24.21%	22.14%	20.42%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	1.0565 (0.3077)	1.0511 (0.2458)	1.0376 (0.2181)	1.0389 (0.2040)
	16.32%	15.78%	14.40%	14.56%

3. *Il C_s di Wright:*

	n = 50	n = 10	n = 150	n = 200
C_s	0.1169 (0.0622)	0.0704 (0.0265)	0.0520 (0.0160)	0.0437 (0.0125)
	77.63%	82.28%	84.12%	84.95%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.9141 (0.4390)	0.8359 (0.2687)	0.7840 (0.1907)	0.7741 (0.1684)
C_{pk}^{WSD}	0.9276 (0.4895)	0.8435 (0.2958)	0.7873 (0.2078)	0.7768 (0.1830)
C_{pm}^{WSD}	0.9300 (0.3215)	0.8792 (0.2246)	0.8370 (0.1773)	0.8287 (0.1584)
	3.06%	4.03%	9.04%	10.00%

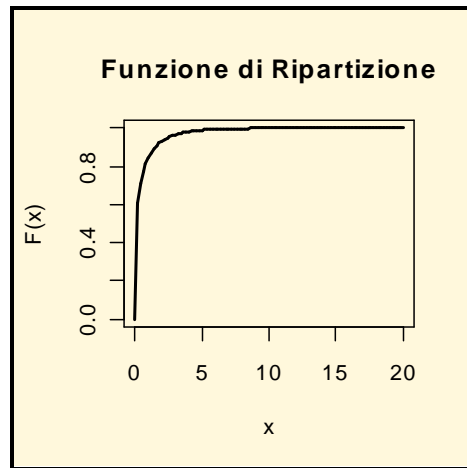
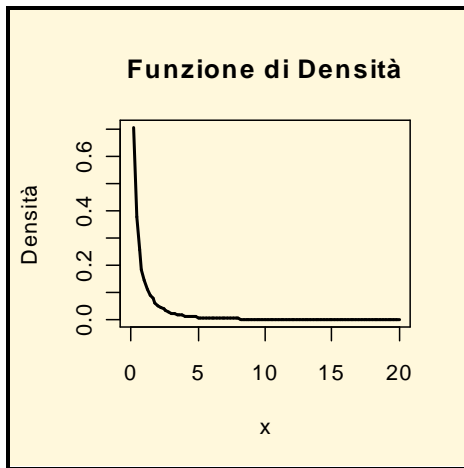
5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	1.1605 (0.4974)	1.0768 (0.3164)	1.0204 (0.2367)	1.0086 (0.2109)
C_{pk}^{WV}	0.7305 (0.3629)	0.6650 (0.2268)	0.6239 (0.1684)	0.6147 (0.1497)
C_{pm}^{WV}	0.7286 (0.3121)	0.6646 (0.2036)	0.6254 (0.1570)	0.6155 (0.1398)
C_{pmk}^{WV}	0.7370 (0.3418)	0.6696 (0.2211)	0.6275 (0.1682)	0.6174 (0.1469)
	18.76%	21.60%	23.25%	23.68%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	1.6986 (23.633)	0.6322 (1.4705)	0.5994 (3.3209)	0.5171 (1.7893)
C_{Npk}	0.7463 (10.393)	0.2783 (0.6495)	0.2637 (1.4557)	0.2276 (0.7880)
C_{Npm}	0.7235 (0.7549)	0.5467 (0.5384)	0.4732 (0.4507)	0.4284 (0.4311)
C_{Npmk}	0.3183 (0.3376)	0.2406 (0.3766)	0.2038 (0.3376)	0.1886 (0.3376)
	42.38%	46.87%	50.70%	55.28%

b) Chi – Quadro ($\chi^2_{0.5}$)



1. Il $C_p, C_{pk}, C_{pm}, C_{pmk}$ con il metodo di Clements:

	n = 50	n = 100	n = 150	n = 200
C_p	1.4965 (0.6542)	1.1515 (0.3932)	1.0190 (0.3143)	0.9474 (0.2718)
C_{pk}	0.8726 (0.3926)	0.6669 (0.2325)	0.5888 (0.1849)	0.5467 (0.1592)
C_{pm}	1.2211 (0.3556)	1.0156 (0.2733)	0.9054 (0.2296)	0.8554 (0.2119)
C_{pmk}	0.8146 (0.3079)	0.6438 (0.2114)	0.5593 (0.1645)	0.5232 (0.1477)
	27.64%	21.46%	19.16%	18.12%

2. Il C_{jpk} :

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.9445 (0.2613)	0.9326 (0.2013)	0.9344 (0.1828)	0.9423 (0.1708)
	8.79%	7.60%	7.78%	8.57%

3. Il C_s di Wright:

	n = 50	n = 100	n = 150	n = 200
C_s	0.1057 (0.0500)	0.0635 (0.0217)	0.0489 (0.1458)	0.0412 (0.0107)
	75.08%	79.30%	80.76%	81.53%

4. Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.8192 (0.3477)	0.7386 (0.2079)	0.7194 (0.1661)	0.7123 (0.1431)
C_{pk}^{WSD}	0.8326 (0.3916)	0.7485 (0.2332)	0.7226 (0.1842)	0.7151 (0.1584)
C_{pm}^{WSD}	0.8151 (0.2568)	0.7610 (0.1755)	0.7471 (0.1466)	0.7417 (0.1280)
	3.42%	10.88%	12.68%	13.35%

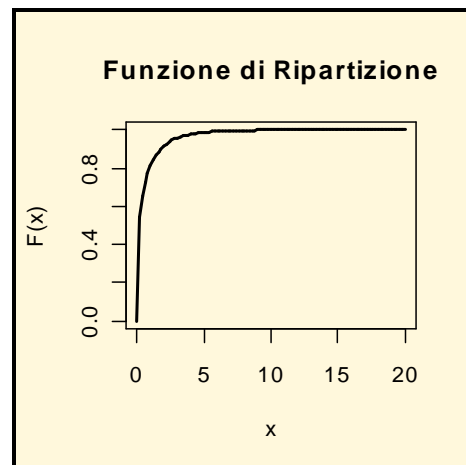
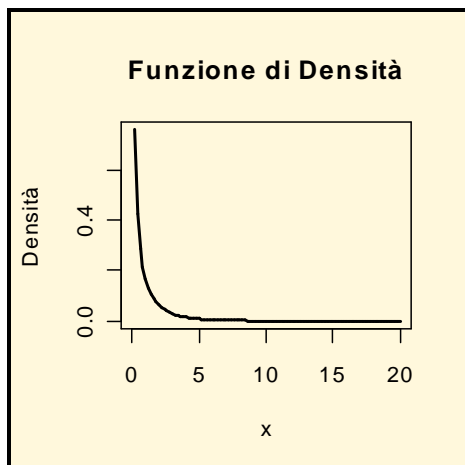
5. Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	1.0418 (0.3956)	0.9555 (0.2452)	0.9378 (0.2005)	0.9301 (0.1744)
C_{pk}^{WV}	0.6735 (0.3044)	0.6034 (0.1846)	0.5888 (0.1508)	0.5821 (0.1302)
C_{pm}^{WV}	0.6705 (0.2616)	0.6046 (0.1672)	0.5897 (0.1380)	0.5824 (0.1194)
C_{pmk}^{WV}	0.6793 (0.2909)	0.6076 (0.1829)	0.5917 (0.1503)	0.5843 (0.1299)
	18.29%	21.32%	22.01%	22.35%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	1.1033 (6.2017)	0.7386 (4.7548)	0.5076 (3.1186)	0.4061 (2.5965)
C_{Npk}	0.4742 (2.6647)	0.3179 (2.0310)	0.2191 (1.3504)	0.1751 (0.1954)
C_{Npm}	0.7043 (0.6501)	0.5155 (0.5062)	0.4053 (0.3098)	0.3752 (0.2956)
C_{Npmk}	0.3028 (0.3376)	0.2223 (0.3376)	0.1748 (0.3376)	0.1618 (0.3376)
	33.37%	40.79%	52.98%	57.70%

c) Chi – Quadro ($\chi^2_{0.6}$)



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.3773 (0.5413)	1.0611 (0.3487)	0.9420 (0.2780)	0.8766 (0.2365)
C_{pk}	0.8238 (0.3381)	0.6283 (0.2136)	0.5558 (0.1687)	0.5161 (0.1426)
C_{pm}	1.1230 (0.3040)	0.9366 (0.2413)	0.8555 (0.2118)	0.8001 (0.1840)
C_{pmk}	0.7690 (0.2761)	0.6069 (0.1969)	0.5421 (0.1571)	0.5007 (0.1332)
	22.97%	19.12%	17.53%	16.44%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.8758 (0.2174)	0.8544 (0.1767)	0.8568 (0.1586)	0.8587 (0.1465)
	4.92%	2.82%	3.02%	3.21%

3. *Il C_s di Wright:*

	n = 50	n = 100	n = 150	n = 200
C_s	0.0998 (0.0436)	0.0592 (0.0197)	0.6709 (0.0130)	0.0385 (0.0094)
	72.67%	76.73%	78.06%	78.80%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.7598 (0.2783)	0.6868 (0.1846)	0.6709 (0.1468)	0.6571 (0.1224)
C_{pk}^{WSD}	0.7730 (0.3170)	0.6930 (0.2099)	0.6756 (0.1659)	0.6594 (0.1378)
C_{pm}^{WSD}	0.7423 (0.2108)	0.6898 (0.1484)	0.6789 (0.1228)	0.6694 (0.1045)
	6.84%	13.66%	15.14%	16.45%

5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	0.9623 (0.3196)	0.8850 (0.2117)	0.8680 (0.1716)	0.8548 (0.1450)
C_{pk}^{WV}	0.6380 (0.2601)	0.5711 (0.1678)	0.5562 (0.1348)	0.5453 (0.1132)
C_{pm}^{WV}	0.6345 (0.2270)	0.5706 (0.1493)	0.5554 (0.1211)	0.5457 (0.1023)
C_{pmk}^{WV}	0.6441 (0.2534)	0.5746 (0.1660)	0.5586 (0.1341)	0.5472 (0.1128)
	17.47%	20.54%	21.27%	21.73%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	1.7327 (20.694)	0.5511 (1.1770)	0.4934 (1.5959)	0.3715 (0.6268)
C_{Npk}	0.7334 (8.8326)	0.2336 (0.5020)	0.2087 (0.6694)	0.1425 (0.2530)
C_{Npm}	0.6728 (0.6326)	0.4755 (0.4490)	0.4092 (0.3310)	0.3470 (0.2261)
C_{Npmk}	0.2844 (0.3376)	0.2014 (0.3376)	0.1733 (0.3376)	0.1328 (0.3376)
	42.37%	46.11%	50.53%	57.80%

6.4. Analisi della distribuzione t di Student (t_n).

A questo punto, studiato come si comportano gli Indici di Capacità per delle distribuzioni asimmetriche vediamo cosa succede per una distribuzione simmetrica, la t di Student. Definita quindi anche questa distribuzione la media del processo e i limiti di specifica, cercherò rispettivamente 3 parametri n in modo che la media della t di Student risulti pari a 3 ed il coefficiente di curtosi vari in ordine decrescente, ottenendo così:

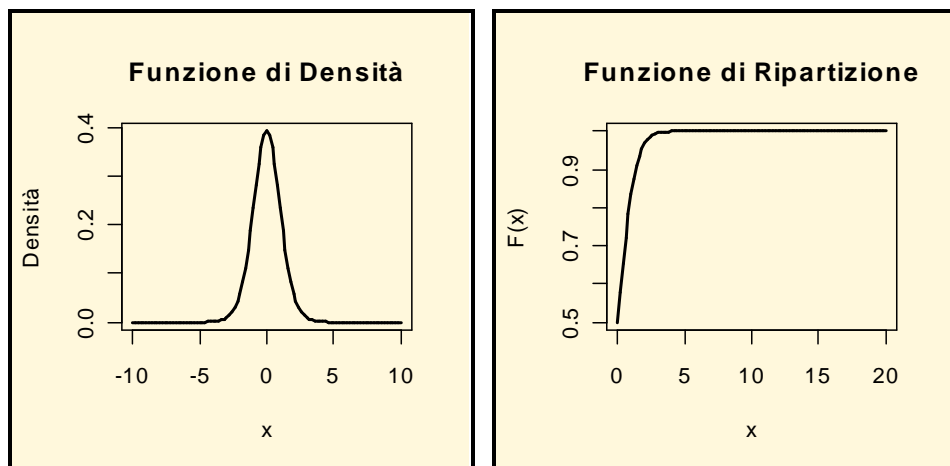
Distribuzione	Media	Varianza	Asimmetri a ($\sqrt{\beta_1}$)	Curtosi (β_2)	NC	C _{pk}
t ₁₅	3	1	0	3.545	0.01%	1.297
t ₂₃	3	1	0	3.315	0.01%	1.297
t ₂₇	3	1	0	3.260	0.01%	1.297

Per ognuno dei tre casi da studiare simulo quindi 1000 campioni di numerosità campionaria 50, 100, 150 e 200 e per ognuno dei 1000 campioni calcolo i vari Indici di Capacità con il metodo di Clements, con il C_{jkp}, con il metodo delle varianze pesate, con il metodo delle deviazioni standard pesate, con il metodo di Wright e infine con il metodo generalizzato.

All'interno di ogni tabella viene sempre rappresentato, per ogni numerosità campionaria, il valore dell'Indice e tra parentesi la sua deviazione standard. Da notare come il valore dell'Indice calcolato sia la media di tutti gli Indici calcolati per i mille campioni e come lo stesso valga anche per la deviazione standard.

Da notare come i coefficienti scelti siano abbastanza alti. Questa scelta è dovuta dal fatto che per la t di Student è impossibile ottenere una distribuzione con varianza pari a 1, dal momento che la sua varianza è pari a n/(n-2). Quindi scegliendo dei parametri abbastanza grandi avvicino la t di Student ad una Normale di media 3 e varianza pari ad 1, non riuscendo mai però a renderla perfettamente uguale.

a) t di Student (t₁₅)



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.2183 (0.2032)	1.0589 (0.1631)	0.9752 (0.1338)	0.9507 (0.1224)
C_{pk}	1.0840 (0.2120)	0.9527 (0.1777)	0.8811 (0.1489)	0.8672 (0.1371)
C_{pm}	1.1898 (0.1925)	1.0485 (0.1594)	0.9695 (0.1316)	0.9470 (0.1213)
C_{pmk}	1.0614 (0.2031)	0.9448 (0.1742)	0.8767 (0.1469)	0.8642 (0.1361)
	16.04%	16.60%	20.76%	21.46%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.6874 (0.0730)	0.6812 (0.0500)	0.6760 (0.0417)	0.6798 (0.0353)
	60.83%	61.57%	62.09%	61.71%

3. *Il C_s di Wright:*

	n = 50	n = 100	n = 150	n = 200
C_s	0.3021 (0.1606)	0.2488 (0.1517)	0.2215 (0.1354)	0.2051 (0.1259)
	100.11%	104.81%	107.54%	109.18%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.8724 (0.1101)	0.8820 (0.0808)	0.8835 (0.0664)	0.8961 (0.0589)
C_{pk}^{WSD}	0.8391 (0.1200)	0.8574 (0.0900)	0.8629 (0.0724)	0.8783 (0.0661)
C_{pm}^{WSD}	0.6649 (0.0731)	0.6610 (0.0512)	0.6567 (0.0418)	0.6615 (0.0364)
	50.15%	49.68%	49.59%	48.49%

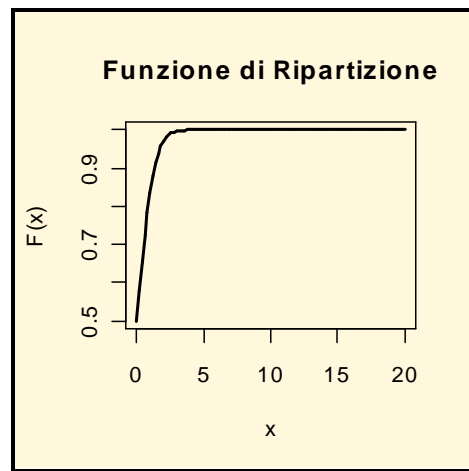
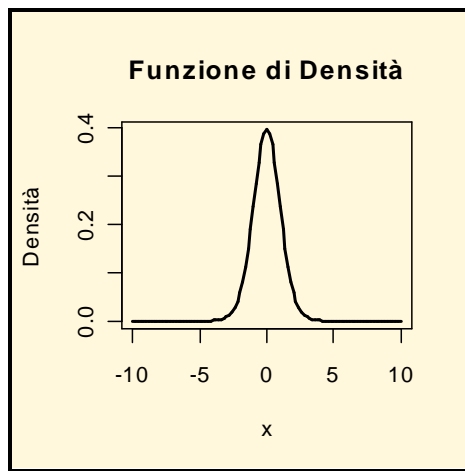
5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	0.9471 (0.1057)	0.9383 (0.0729)	0.9310 (0.0595)	0.9372 (0.0516)
C_{pk}^{WV}	0.9730 (0.1163)	0.8826 (0.0841)	0.8857 (0.0666)	0.8976 (0.0592)
C_{pm}^{WV}	0.8872 (0.1119)	0.8921 (0.0805)	0.8933 (0.0.63)	0.9039 (0.0563)
C_{pmk}^{WV}	0.8740 (0.1160)	0.8829 (0.0847)	0.8858 (0.0670)	0.8977 (0.0595)
	40.15%	39.79%	39.79%	38.78%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.4629 (0.2248)	0.4128 (0.0580)	0.3887 (0.0505)	0.3722 (0.0482)
C_{Npk}	0.2277 (0.1140)	0.2047 (0.0289)	0.1928 (0.0260)	0.1852 (0.0249)
C_{Npm}	0.4624 (0.2222)	0.4126 (0.0579)	0.3886 (0.0506)	0.8961 (0.0482)
C_{Npmk}	0.2274 (0.3376)	0.2046 (0.3376)	0.1928 (0.3376)	0.6615 (0.3376)
	95.32%	98.82%	100.62%	101.83%

b) t di Student (t_{23})



1. Il $C_p, C_{pk}, C_{pm}, C_{pmk}$ con il metodo di Clements:

	n = 50	n = 100	n = 150	n = 200
C_p	1.2676 (0.2045)	1.1231 (0.1538)	1.0512 (0.1389)	1.0075 (0.1170)
C_{pk}	1.1400 (0.2116)	1.0199 (0.1673)	0.9605 (0.1535)	0.9209 (0.1344)
C_{pm}	1.2381 (0.1943)	1.1112 (0.1504)	1.0456 (0.1368)	0.9974 (0.1158)
C_{pmk}	1.1169 (0.2027)	1.0107 (0.1639)	0.9556 (0.1516)	0.9175 (0.1331)
	7.68%	13.78%	16.79%	18.77%

2. Il C_{jkr} :

	n = 50	n = 100	n = 150	n = 200
C_{jkr}	0.7013 (0.0722)	0.6944 (0.0531)	0.6924 (0.0408)	0.6889 (0.0369)
	59.56%	60.25%	60.45%	60.80%

3. Il C_s di Wright:

	n = 50	n = 100	n = 150	n = 200
C_s	0.3378 (0.1808)	0.2668 (0.1438)	0.2361 (0.1399)	0.2269 (0.1422)
	96.21%	103.03%	106.08%	107%

4. Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.8976 (0.1079)	0.9072 (0.0815)	0.9158 (0.0649)	0.9168 (0.0601)
C_{pk}^{WSD}	0.8646 (0.1169)	0.8821 (0.0904)	0.8961 (0.0718)	0.8994 (0.0672)
C_{pm}^{WSD}	0.6826 (0.0728)	0.6797 (0.0562)	0.6782 (0.0417)	0.6756 (0.0374)
	48.20%	47.39%	46.69%	46.63%

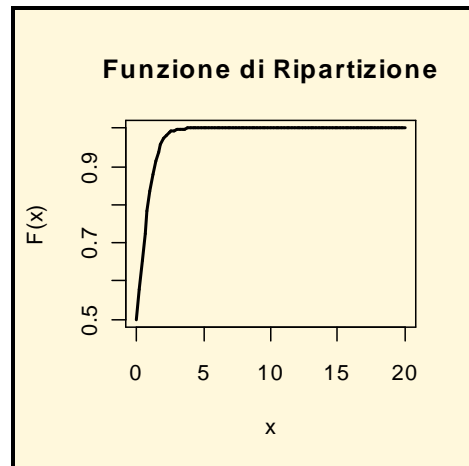
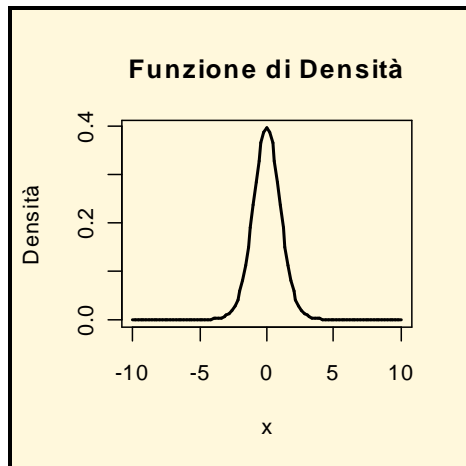
5. Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	0.9723 (.1065)	0.9650 (0.0752)	0.9613 (0.0594)	0.9572 (0.0530)
C_{pk}^{WV}	0.9010 (0.1139)	0.9108 (0.0815)	0.9166 (0.0663)	0.9190 (0.0595)
C_{pm}^{WV}	0.9115 (0.1099)	0.9203 (0.0787)	0.9237 (0.0637)	0.9252 (0.0574)
C_{pmk}^{WV}	0.9021 (0.1130)	0.9108 (0.0818)	0.9171 (0.0665)	0.9192 (0.0598)
	37.43%	37.02%	36.72%	36.67%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.4723 (0.3322)	0.4189 (0.0608)	0.3918 (0.0415)	0.3791 (0.0352)
C_{Npk}	0.2316 (0.1593)	0.2076 (0.0315)	0.1948 (0.0213)	0.1890 (0.0180)
C_{Npm}	0.4700 (0.2756)	0.4187 (0.0605)	0.3917 (0.0415)	0.3790 (0.0352)
C_{Npmk}	0.2305 (0.3376)	0.2075 (0.3376)	0.3917 (0.3376)	0.1890 (0.3376)
	94.58%	98.37%	100.03%	101.29%

c) t di Student (t_{27})



1. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo di Clements:*

	n = 50	n = 100	n = 150	n = 200
C_p	1.2894 (0.2042)	1.1346 (0.1516)	1.0629 (0.1314)	1.0150 (0.1200)
C_{pk}	1.1574 (0.2171)	1.0314 (0.1659)	0.9732 (0.1469)	0.9343 (0.1340)
C_{pm}	1.2576 (0.1929)	1.1215 (0.1464)	1.0565 (0.1296)	1.0111 (0.1182)
C_{pmk}	1.1328 (0.2080)	1.0212 (0.1610)	0.9677 (0.1451)	0.9307 (0.1324)
	6.79%	13.20%	16.15%	18.11%

2. *Il C_{jpk} :*

	n = 50	n = 100	n = 150	n = 200
C_{jpk}	0.7053 (0.0713)	0.6968 (0.0503)	0.6939 (0.0391)	0.6229 (0.0367)
	59.16%	60.01%	60.30%	60.40%

3. *Il C_s di Wright:*

	n = 50	n = 100	n = 150	n = 200
C_s	0.3303 (0.1717)	0.2732 (0.1634)	0.2436 (0.1451)	0.2337 (0.1427)
	96.66%	102.37%	105.33%	106.32%

4. *Il C_p , C_{pk} , C_{pm} con il metodo della deviazione standard pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WSD}	0.9060 (0.1089)	0.9126 (0.0804)	0.9187 (0.0643)	0.9238 (0.0583)
C_{pk}^{WSD}	0.8732 (0.1189)	0.8881 (0.0894)	0.8984 (0.0718)	0.9065 (0.0649)
C_{pm}^{WSD}	0.6881 (0.0722)	0.6826 (0.0510)	0.6013 (0.0408)	0.6805 (0.0370)
	47.45%	46.91%	46.41%	46%

5. *Il C_p , C_{pk} , C_{pm} , C_{pmk} con il metodo della varianza pesata:*

	n = 50	n = 100	n = 150	n = 200
C_p^{WV}	0.9803 (0.1051)	0.9688 (0.0728)	0.9657 (0.0582)	0.9640 (0.0525)
C_{pk}^{WV}	0.9071 (0.1125)	0.9159 (0.0809)	0.9208 (0.0645)	0.9265 (0.0565)
C_{pm}^{WV}	0.9206 (0.1084)	0.9236 (0.0775)	0.9274 (0.0624)	0.9316 (0.0547)
C_{pmk}^{WV}	0.9083 (0.1126)	0.9162 (0.0811)	0.9211 (0.0647)	0.9267 (0.0567)
	36.78%	36.58%	36.32%	35.94%

6. Il $C_{Np}(u, v)$ generalizzato:

	n = 50	n = 100	n = 150	n = 200
C_{Np}	0.4599 (0.0708)	0.4178 (0.0492)	0.3930 (0.0408)	0.3824 (0.0394)
C_{Npk}	0.2259 (0.0362)	0.2064 (0.0250)	0.1954 (0.0212)	0.1906 (0.0201)
C_{Npm}	0.4594 (0.0704)	0.4177 (0.0491)	0.3930 (0.0408)	0.3824 (0.0393)
C_{Npmk}	0.2257 (0.3376)	0.2063 (0.3376)	0.1953 (0.3376)	0.1906 (0.3376)
	95.42%	98.49%	100.27%	101.04%

6.5. Risultati finali e analisi finale.

Come sottolineato nel metodo di ricerca, la mia analisi si concentra su due differenti tipi di distribuzioni, le distribuzioni asimmetriche e simmetriche. Questa mia scelta è dovuta al fatto che voglio ricavare quale sono i metodi più corretti e adatti per i due tipi di distribuzioni. Per quanto riguarda le asimmetriche ho preso in considerazione la Gamma e la Chi – Quadro, mentre per le simmetriche la t di Student.

Cominciamo a questo punto con l'osservare le distribuzioni asimmetriche e i loro risultati. Tra questi risultati possiamo ancora fare una distinzione, tra i metodi che risultano essere più precisi e quindi adatti alla stima della Capacità del processo e quelli che risultano essere meno adatti perché hanno una differenza media percentuale dal valore di riferimento troppo elevata.

Cominciamo con i metodi meno adatti per le distribuzioni asimmetriche e ricaviamo che questi sono, sia per la Gamma che per la Chi – Quadro, il metodo di Clements, il metodo del C_s di Wright e il metodo Generalizzato, il metodo del $C_{j_{kp}}$ ed il metodo delle varianze pesate. Per quanto riguarda il metodo di Clements applicato alla Gamma osserviamo come sia il C_p che il C_{pm} siano, al diminuire del coefficiente di asimmetria e curtosi e all'aumentare della numerosità campionaria, sovrastimati e superiori del C_{pk} equivalente mentre il C_{pk} e il C_{pmk} siano sempre rispetto al C_{pk} equivalente, al diminuire del coefficiente di asimmetria e curtosi, prima sovrastimati per $n = 50$ e poi sottostimati per n maggiore di 50. Da notare anche come i salti tra il (C_p, C_{pm}) e il (C_{pk}, C_{pmk}) siano molto elevati e come la differenza media percentuale che si incontra con l'applicazione di questo metodo per la Gamma sia, all'aumentare di n e al diminuire dell'asimmetria e curtosi, intorno al 17%, una differenza percentuale media abbastanza alta per poter risultare una buona stima rispetto al C_{pk} equivalente.

Per quanto riguarda la Chi – Quadro il risultato mi porta alle stesse conclusioni della Gamma. Infatti per questa distribuzione osserviamo come sia il C_p che il C_{pm} siano, al diminuire del coefficiente di asimmetria e curtosi e all'aumentare della numerosità campionaria, sovrastimati e superiore del C_{pk} equivalente mentre il C_{pk} e il C_{pmk} siano sempre sottostimati rispetto al C_{pk} effettivo. Anche per la Chi - Quadro la differenza percentuale media è abbastanza alta, intorno al 20%, una differenza media abbastanza elevata per poter risultare una buona stima rispetto al C_{pk} equivalente.

Questa sostanziale differenza quindi è da attribuire ad una situazione particolare che in queste distribuzioni viene a crearsi, per la precisione un'asimmetria ed una curtosi molto elevata. Infatti il metodo di Clements è particolarmente adatto per le distribuzioni simmetriche o limitatamente asimmetriche, con un coefficiente di asimmetria intorno al 2%, portando quindi le sue stime in questo caso a delle errate e non precise conclusioni sulla Capacità del processo.

Ora prendiamo in considerazione il secondo metodo meno adatto, il metodo del C_s di Wright. Osservando la tabella con i valori di questo indice ricaviamo come, sia per la Gamma che per la Chi – Quadro, il C_s sia molto sottostimato ed inferiore rispetto al C_{pk} equivalente. La sua differenza percentuale media è pari a 83% per la Gamma e 79% per la Chi – Quadro ed aumenta all'aumentare della numerosità campionaria, una differenza media molto elevato per poter risultare la stima corretta e precisa. Questa differenza è così elevata a causa dell'uso corretto che deve essere fatto di C_s , infatti in generale il C_s fornisce un approccio statistico per l'analisi della capacità del processo sotto tre condizioni: $USL - \mu \neq \mu - LSL$, $\mu \neq T$, $\sqrt{\beta_1} \neq 0$. Dal momento che noi abbiamo assunto tre condizioni differenti da queste, e precisamente $USL - \mu = \mu - LSL$, $\mu = T$, $\sqrt{\beta_1} \neq 0$, la stima di conseguenza non sarà buona ma sottostimata.

Altro approccio poco adatto è quello del metodo generalizzato, metodo per l'appunto troppo generale per poter ricavare dei valori vicini al C_{pk} equivalente. Infatti, rispettivamente per la Gamma e per la Chi – Quadro, notiamo come ci sia una differenza percentuale media che aumenta all'aumentare della numerosità campionaria e che è intorno al 60% e 47%. Differenza dovuta fondamentalmente al fatto che, per il caso di distribuzioni Non-Normali, le stime diventano molto instabili a causa della distribuzione delle varianze campionarie, s^2 , stime che sono sensibili alla deviazione dalla Normalità.

Prendendo poi in considerazione il $C_{j_{kp}}$, osserviamo come per la Gamma questo sia, al diminuire dell'asimmetria e curtosi, prima sovrastimato e superiore al C_{pk} equivalente e poi inferiore. Cosa molto particolare è anche l'aumento considerevole che assume la differenza media percentuale al diminuire del coefficiente di asimmetria, infatti da una differenza media percentuale intorno all'1.5% si passa all'8% per finire con un 18%. Accade l'opposto per la Chi – Quadro, infatti al diminuire dell'asimmetria e curtosi, il $C_{j_{kp}}$ è sempre superiore al valore nominale, con una differenza percentuale media che passa dal 15% al 7.5% per finire con un 3.5%. Da notare sia nel caso della Chi – Quadro che della Gamma come il $C_{j_{kp}}$ all'aumentare della numerosità campionaria rimanga per lo più attorno allo stesso valore variando di poco. Cosa molto singolare che quindi rende questo metodo poco attendibile per il calcolo della Capacità delle distribuzioni asimmetriche sono questi salti elevati che colpiscono il $C_{j_{kp}}$ al diminuire dell'asimmetria.

E per finire l'ultimo metodo poco adatto, quello delle varianze pesate. Da osservare come al diminuire del coefficiente di asimmetria e curtosi anche la differenza media percentuale diminuisca, per la Gamma da circa il 18% al 9%, mentre per la Chi – Quadro rimanga per lo più costante, attorno al 20%. A casua di questa discontinuità e alta percentuale quindi possiamo considerare questo metodo poco adatto per le distribuzioni Non – Normali e asimmetriche.

Consideriamo ora il metodo che risulta essere più adatto per le distribuzioni asimmetriche, quello della deviazione standard pesata. Cominciamo osservando cosa accade alla

differenza percentuale media delle due distribuzioni quando l'asimmetria diminuisce. Per la Gamma questa risulta rimanere intorno al 12% mentre per la Chi – Quadro passa da circa il 6% al 12%. Da un'analisi più dettagliata di questo metodo ricaviamo anche come, al diminuire dei coefficienti di asimmetria e curtosi, sia per la Gamma che per la Chi – Quadro, gli Indici sottostimino il C_{pk} equivalente, proprio come ci aspettavamo dalla teoria studiata nel Capitolo 4 (pag 82 – 83) e come può essere osservato da una analisi grafica (vedere Appendice a). Altra osservazione su questo metodo è la numerosità campionaria e quale risulta essere la migliore. Notiamo infatti come, soprattutto per la Chi – Quadro, ci sia un salto rilevante tra $n = 50$ e $n = 100$ mentre poi per n superiore a 100 la differenza percentuale rimanga per lo più costante per tutte le distribuzioni. Possiamo quindi concludere che $n = 100$ potrebbe essere la numerosità campionaria più adatta per uno studio sulla Capacità di un processo produttivo.

Analizziamo infine la distribuzione simmetrica, la t di Student. Osserviamo come l'unico metodo adatto sia quello di Clements e come gli altri siano poco adatti dal momento che la loro differenza percentuale è molto alta. Infatti al diminuire della curtosi, la differenza percentuale media del $C_{j_{kp}}$ è intorno al 60%, del C_s intorno al 100%, del metodo delle deviazioni standard pesate intorno al 49%, del metodo della varianza pesata intorno al 40% e del metodo generalizzato intorno al 98%. Questo poca adattabilità è dovuta fondamentalmente al fatto che tutti questi metodi sono stati studiati e alcuni risultati adatti per le distribuzioni Non – Normali ma asimmetriche.

Anche il metodo di Clements è stato studiato, come visto precedentemente, per le distribuzioni asimmetriche, risultando però più adatto per le distribuzioni simmetriche come viene osservato dall'analisi della t di Student.

Infatti da questa analisi ricaviamo come, al diminuire del coefficiente di curtosi, la differenza media percentuale passi da circa il 18% a circa il 15%. Anche se sembrerebbe una differenza abbastanza elevata per classificare come adatto il metodo, lo riteniamo il più adatto dal momento che è quello che si avvicina maggiormente al C_{pk} equivalente, notando comunque come ci sia un margine di errore superiore al 10% (Appendice b). Possiamo anche cercare qual'è la numerosità campionaria preferibile per un campionamento. Secondo me, visto che, come per il metodo della deviazione standard per le distribuzioni asimmetriche, notiamo un salto rilevante tra $n = 50$ e $n = 100$, possiamo ritenere che $n = 100$ sia la numerosità campionaria da preferire, anche perché al diminuire della curtosi, la differenza percentuale media si stabilizza intorno al 13%.

CONCLUSIONI

Dopo questa analisi finale dettagliata derivante dalla simulazione posso concludere quindi che il metodo più adatto per le distribuzioni Non – Normali e asimmetriche è quello delle deviazioni standard pesate mentre per le distribuzioni Non – Normali ma simmetriche è quello di Clements.

Da sottolineare però come questi metodi risultino adatti sotto delle condizioni molto particolari e ben precise, infatti si prende in considerazione un processo la cui media è centrata all'interno dei limiti di specifica e il valore obiettivo combacia con la stessa media.

Altra annotazione importante è che la differenza media percentuale calcolata rispetto al C_{pk} equivalente, sia per il metodo delle deviazioni standard sia per il metodo di Clements, è intorno al 12% - 13%, una percentuale abbastanza grande anche se minore rispetto alle altre per poter affermare che si sta stimando con precisione la vera Capacità del processo. Dobbiamo quindi ricordare quando si andrà a stimare la Capacità di un processo, e di conseguenza i relativi Indici di Capacità, come si potrebbe commettere un errore di stima intorno a questa percentuale.

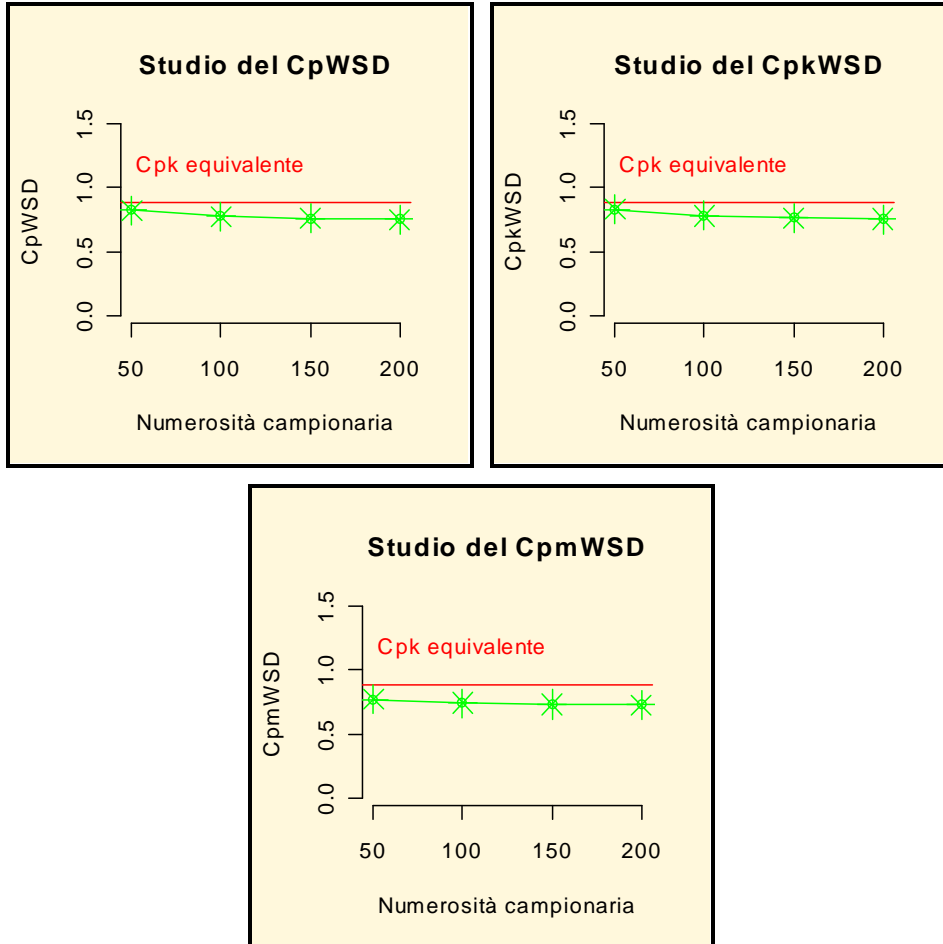
Questa considerazione comunque non macchia l'importanza che hanno avuto e che hanno tuttora gli Indici di Capacità, mettendo anche in risalto però le difficoltà che si incontrano nel loro uso, le particolari condizioni che devono essere rispettate per avere una corretta interpretazione.

Queste difficoltà potrebbero quindi sfiduciare la ricerca del migliore metodo da applicare da parte di chi li deve mettere in pratica, ma queste difficoltà possono essere superate, come afferma Bunker, dall'intelligenza, dalla velocità e dalla forza. Tutte caratteristiche che portano a dei benefici, benefici che soprattutto si hanno quando i processi e gli Indici di Capacità vengono capiti e percepiti, quando chi li usa ha una buona conoscenza della teoria statistica, ottenendo così il massimo risultato: la soddisfazione del cliente.

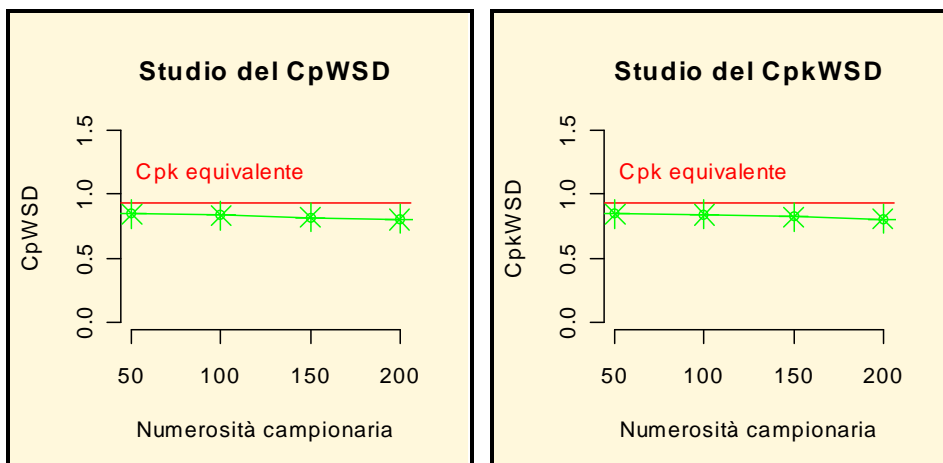
APPENDICE

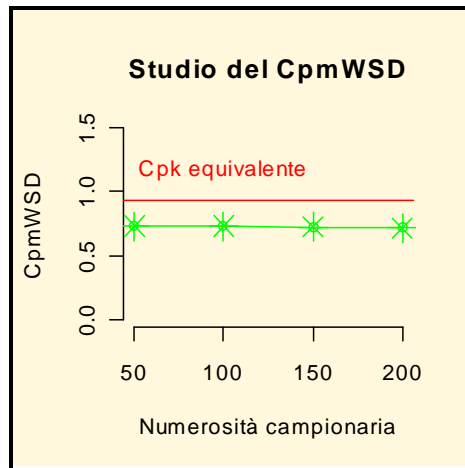
a) Grafici del metodo delle deviazioni standard applicato alla Gamma e Chi – Quadro al diminuire del coefficiente di asimmetria e curtosi:

❖ **Gamma(Γ (0.445, 0.667)).**

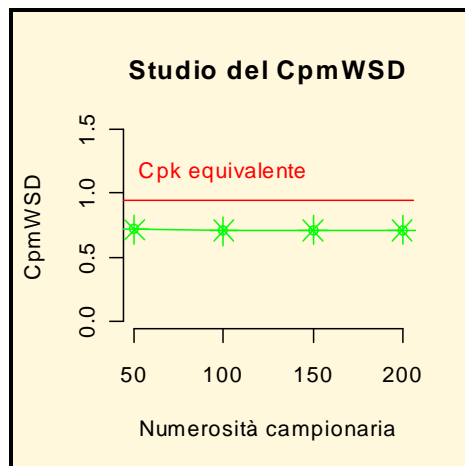
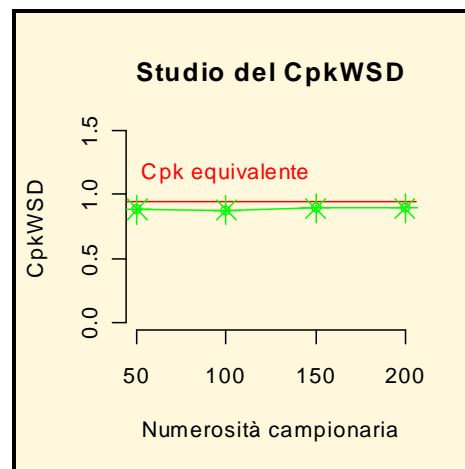
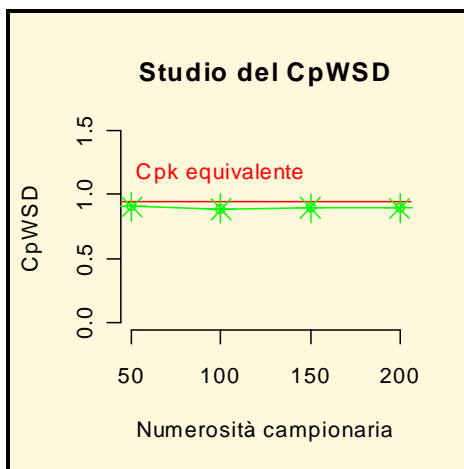


❖ **Gamma(Γ (1, 1)).**

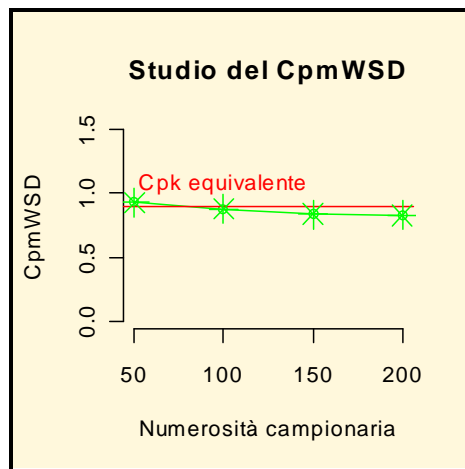
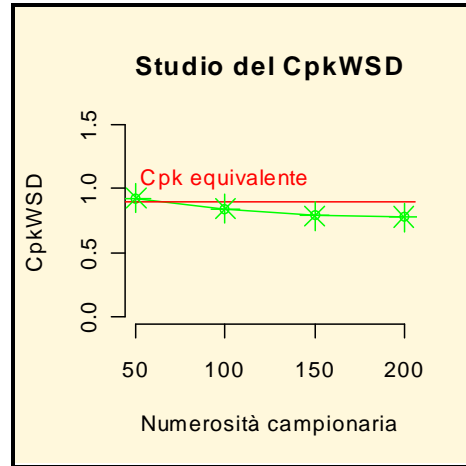
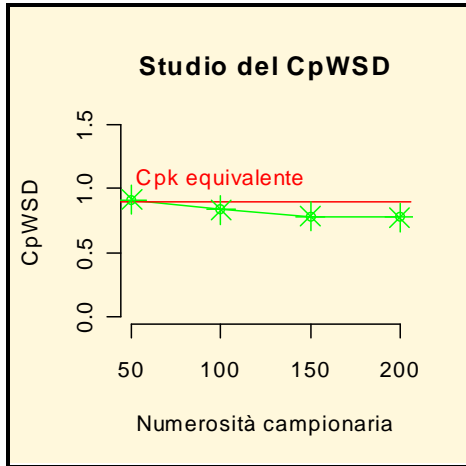




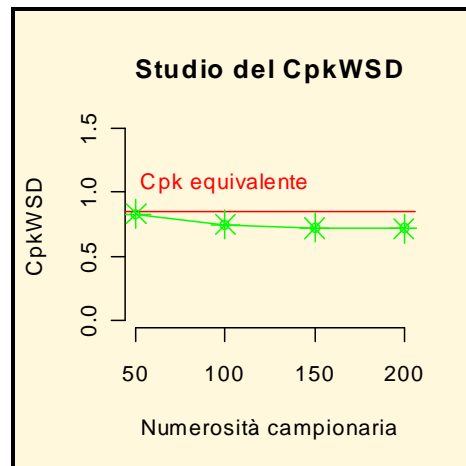
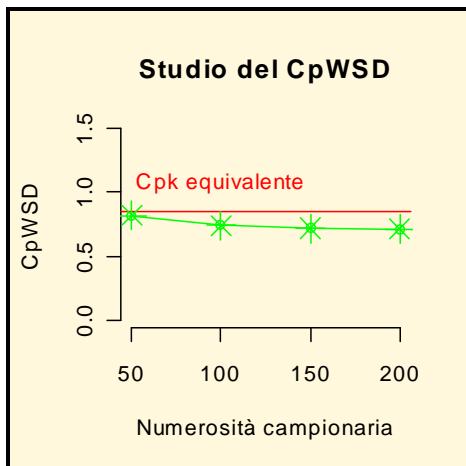
❖ **Gamma($\Gamma(4, 2)$).**

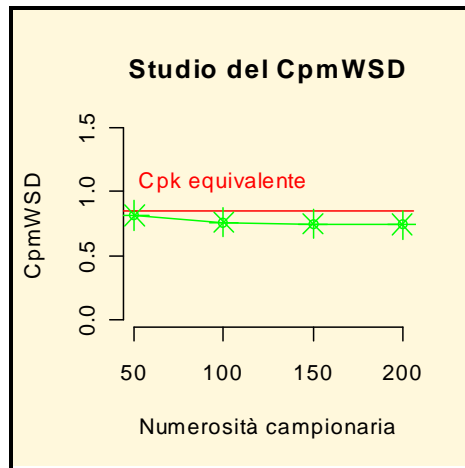


▪ Chi – Quadro ($\chi^2_{0.4}$).

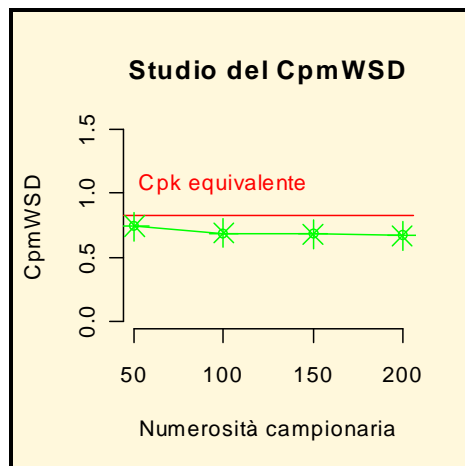
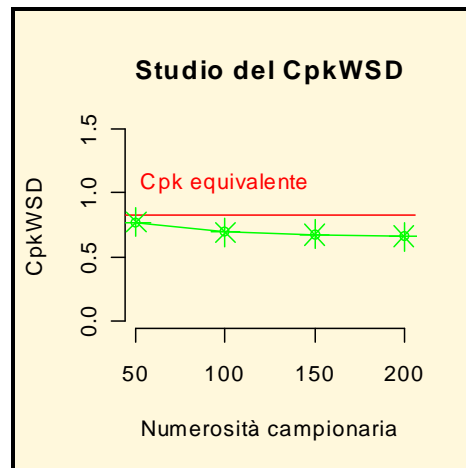
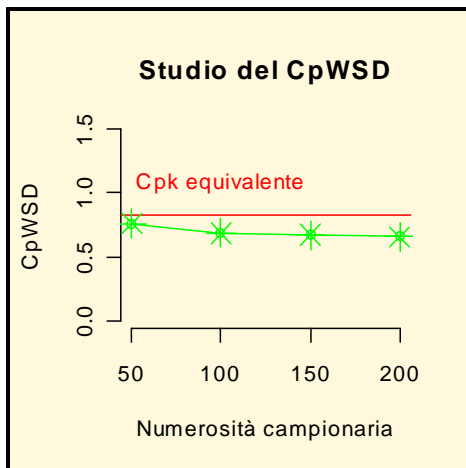


▪ Chi – Quadro ($\chi^2_{0.5}$).



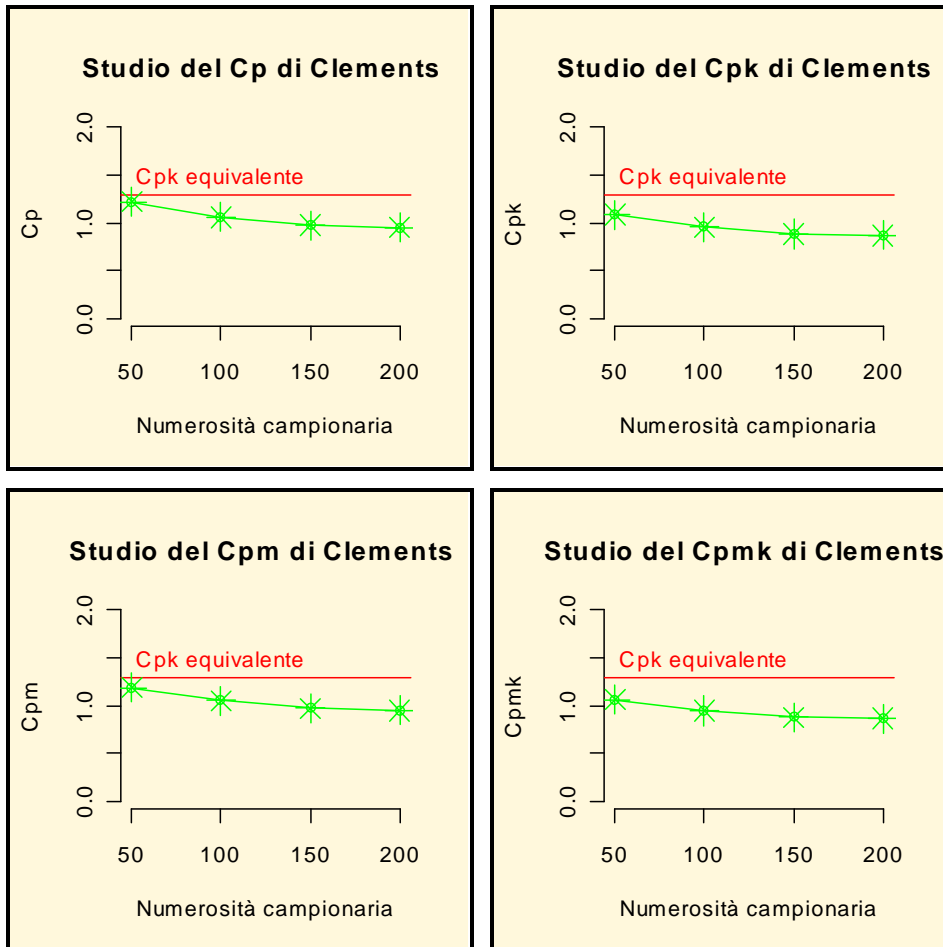


- Chi - Quadro ($\chi^2_{0,6}$).

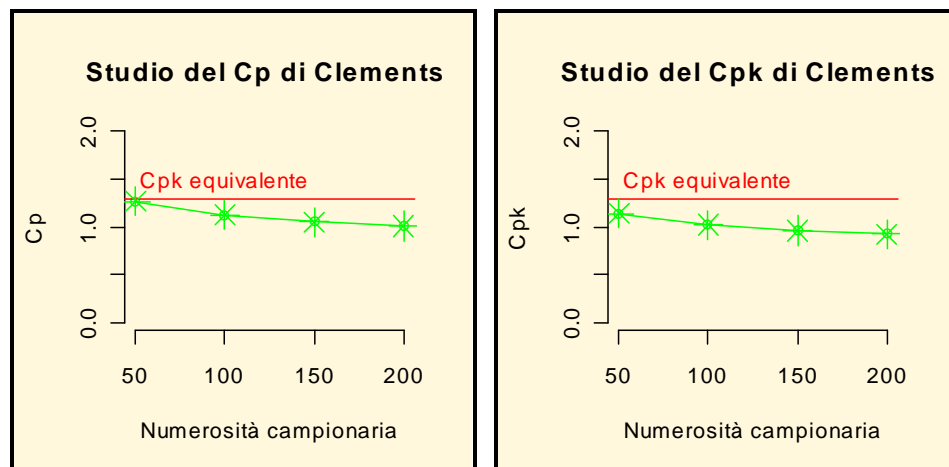


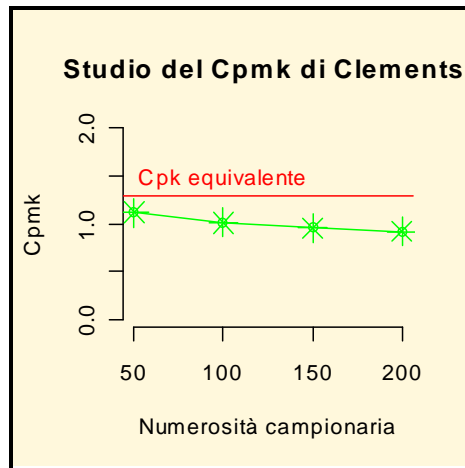
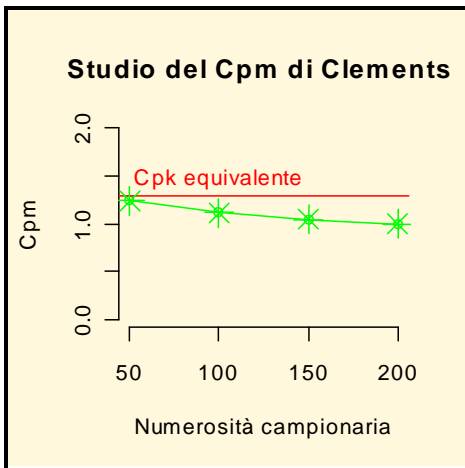
b) Grafici del metodo delle deviazioni standard applicato alla t di Student al diminuire del coefficiente di asimmetria e curtosi:

- **t di Student (t_{15}).**

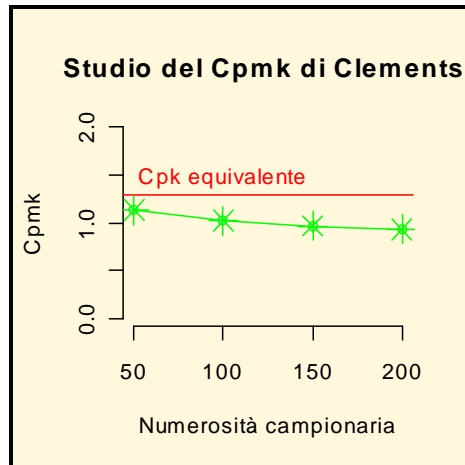
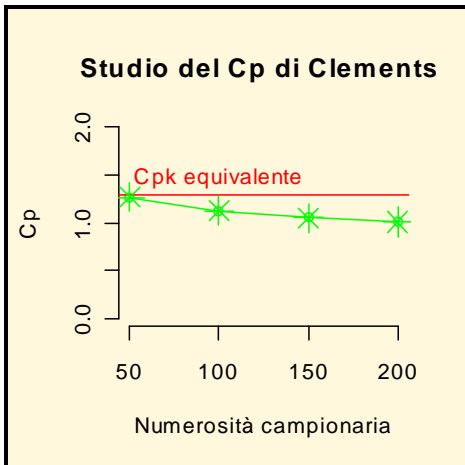
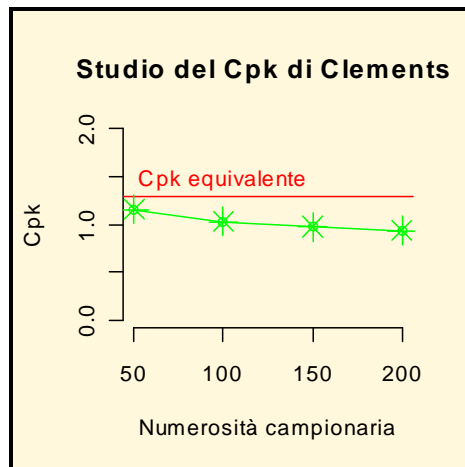
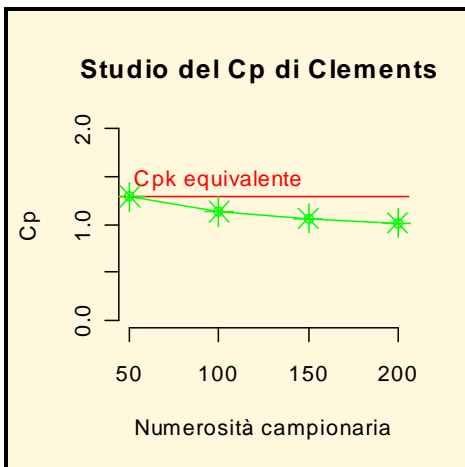


- **t di Student (t_{23}).**





• **t di Student (t_{27}).**



BIBLIOGRAFIA

- Process Capability Indices in Theory and in Practice (Kotz – Lowelace)
- Process Capability Indices (Kotz – Johnson)
- Process Capability Indices – A Review 1992-2000 (Kotz – Johnson, Journal of Quality Technology, Vol. 34, No. 1, pag. 1-19, January 2002)
- Process Capability Studies: What is a Process Capability Studies? (Gunter, Quality Progress, pag. 97-101, February 1991)
- AT&T Statistical Quality Control Handbook (Charlotte, NC: Delmar Printing Company, 1995, pag.45-46)
- AT&T Statistical Quality Control Handbook (Charlotte, NC: Delmar Printing Company, 1995, pag.45)
- Understanding Process Capability Indices”, Stefan Steiner, Bovas Abrham e Jock Mackay, Institute for Improvement of Quali y and Productivity Departament of Statistics and Actual Science University of Waterloo
- Burke RJ, Davis RD and Kamisky FC (1991). Responding to statistical terrorism in quality control. Paper presented to the 47th Annual Conference, Rochester Srction of the American Society for Quality control, 21 march 1991
- The Use and Abuse of C_{pk} (Gunter, Quality Progress, pag. 72-87, January 1989)
- Using Performance Indices to Monitor Production Processes (Paul F. McCoy, Quality Progress, pag. 49-95, February 1991)
- Analysing the Capability of a Process (Montgomery, 1997)
- Understanding Process Capability Indices (Stefan Steiner, Bovas Abrham e Jock Mackay, Institute for Improvement of Quali y and Productivity Departament of Statistics and Actual Science University of Waterloo)

- Responding to statistical terrorism in quality control (Burke RJ, Davis RD and Kamisky FC, Paper presented to the 47th Annual Conference, Rochester Section of the American Society for Quality control, 21 march 1991)
- Discussion (Fred Spring, Smiley Cheng and Anthony Yeung, Bartholomew Leung, Journal of Quality Technology, Vol. 1, No. 1, pag. 23-27, January 2002)
- Incapability Index with Asymmetric Tolerances (Chen, Statistica Sinica, Vol. 8, pag. 253-262, 1998)
- Non–Normal Distributions in the Real Word (Thomas Pyzdek, www.qualitydigest.com)
- On the Sampling Distribution of Clements Capability Index (Wagner Borges, Linda Lee Ho, Communications Statistics – Simulations, Vol.29, No.1, pag. 131-138, 2000)
- Process Capability Calculations for Non – Normal Distributions (John A. Clements, Quality Progress, pag95 -101, September 1989)
- An application of Non – Normal Process Capability Indices (Chen and Pearn, Quality and Reliability Engineering International, Vol. 13, pag. 355-360, June 1997)
- Abstract of Capability Indices for Non - Normal Data (McCormmack, Harris, Arnon, Spagon, Progress Engineering, Vol.12, pag. 489-495, June 2000)
- Abstract of Process Capability Indices for Non – Normal Distributions (Somerville, Montgomery, Progress Engineering, December 1996)
- Abstract of Using Johnson Curves to describe Non – Normal Process Data (Farnum, Progress Engineering, December 1996)

- A Smooth Nonparametric Approach to Process Capability (Polansky, Quality and Reliability Engineering International, vol. 14, pag. 43-48, November 1997)
- Computing Process Capability Indices for Non – Normal Data (Tang and Than, Quality and Reliability Engineering International, Vol.15, February 1999, pag. 339-353)
- A Weighted Variance Capability Index for General Non – Normal Processes (Wu, Swain, Farrington, Messimer, Quality and Reliability Engineering International, Vol. 15, pag. 397-402, March 1999)
- A Monte Carlo Comparison of Capability Indices when Processes are not Normally Distributed (Wu, Swain, Quality and Reliability Engineering International, Vol. 17, pag. 219-231, January 2001)
- Assessment of a Capability Index Sensitive to Skewness (Nhar, Hubele, Zimmer, Quality and Reliability Engineering International, Vol. 17, pag. 233-241, January 2001)
- Process Capability Indices for Skewed Populations (Chang, Choi, Bai, Quality and Reliability Engineering International, Vol. 18, pag. 383-393, March 2002)
- A Process Capability Index Sensitive to Skewness (Peter A. Wright, School of Mathematical Sciences, University of Technology)
- Discussion (Normal Hubele, Journal of Quality Technology, Vol. 34, No. 1, pag. 38-39, January 2002)
- An Algorithm for Fitting Johnson Transformation to Non – Normal Data (Polansky, Chou, Mason, Journal of Quality Technology, July 1999)
- The Johnson System: Selection and Parameter Estimation (Slikfer and Shapiro, Technometrics, May 1980)

- On Using Percentiles to Fit Data by a Johnson Distribution (Chou, Turner, Henson, Meyer, Chen, Communications, Communications Statistics – Simulations, Vol. 23, No. 32, , pag. 341-3541994)