



**UNIVERSITA' DEGLI STUDI DI PADOVA**

**FACOLTA' DI SCIENZE STATISTICHE**

**CORSO DI LAUREA IN STATISTICA E GESTIONE**

**DELLE IMPRESE**

**TESI DI LAUREA**

**CONFRONTO TRA DUE CARTE EWMA**

**PER DATI NON NORMALI**

**Relatore: Prof.ssa CAPIZZI GIOVANNA**

**Laureando: FASOLATO NICOLA**

**Matricola: 498813**

**ANNO ACCADEMICO 2005/2006**

# INDICE

## CAPITOLO 1

### IL CONTROLLO STATISTICO DI PROCESSO

1.1 Introduzione.....	3
1.2 Le carte di controllo.....	6
1.3 Componenti delle carte di controllo.....	7
1.4 Le carte per variabili e carte per attributi.....	10
1.5 La carta Shewhart.....	14
1.6 La carta EWMA.....	15
1.7 Le carte di controllo non parametriche.....	20

## CAPITOLO 2

### CARTA EWMA BASATA SUI RANGHI

2.1 Introduzione.....	22
2.2 ARL per processi in controllo.....	23
2.3 ARL per processi fuori controllo.....	24
2.4 Conclusioni.....	25

## **CAPITOLO 3**

### **CARTA EWMA PER DISTRIBUZIONI DI WEIBULL**

<b>3.1 Introduzione.....</b>	<b>26</b>
<b>3.2 Distribuzione di Weibull.....</b>	<b>26</b>
<b>3.3 Carta EWMA per monitorare la media di un processo caratterizzato da distribuzione di Weibull.....</b>	<b>30</b>

## **CAPITOLO 4**

<b>4.1 Introduzione.....</b>	<b>34</b>
<b>4.2 Carte Ewma con dati caratterizzati da distribuzione di Weibull.....</b>	<b>34</b>
<b>4.3 Carte EWMA con dati non normali.....</b>	<b>39</b>
<b>4.4 Carte EWMA con dati normali.....</b>	<b>43</b>
<b>4.5 Commenti.....</b>	<b>47</b>

<b>BIBLIOGRAFIA.....</b>	<b>48</b>
--------------------------	-----------

# CAPITOLO 1

## IL CONTROLLO STATISTICO DI PROCESSO

### 1.1 Introduzione

L'ascolto dell'ambiente esterno e l'individuazione dell'efficienza, presi come fattori necessari per fornire prodotti e servizi di qualità, ha comportato nelle aziende un'attenzione sempre maggiore verso il controllo dei processi.

Ogni azienda, infatti, erogando prodotti o servizi può essere studiata come un insieme di processi i quali presi nella loro globalità consentono di analizzare nel migliore dei modi, gli sforzi finalizzati al miglioramento di efficacia, efficienza, flessibilità e adattabilità nelle attività aziendali svolte.

La maggior attenzione verso il controllo dei processi è sottolineata anche dalla norma ISO 9001 del 2000 che ha legato il concetto di processo a quello di miglioramento continuo.

La norma (*"L'organizzazione deve adottare adeguati metodi per misurare e monitorare quei requisiti necessari per rispondere ai requisiti del cliente.*

*Questi devono confermare la continua capacità dei singoli processi per ottemperare le relative finalità"* ISO 9001:2000 capitolo 8 paragrafo 2.3)

richiede dunque all'azienda di attivare studi e controlli dei propri processi per assicurare non più soltanto il mantenimento di determinati livelli di qualità ma anche il loro miglioramento continuo.

I principi di gestione cui occorre uniformare l'azienda per muoversi secondo quest'ottica di processo e tendere così al miglioramento continuo sostanzialmente sono:

- Definire in modo chiaro e lineare il processo e le attività;
- Descrivere gli ambiti del processo e i punti di interfaccia con clienti e fornitori;
- Determinare in modo chiaro e univoco i "responsabili" del processo;
- Stabilire i fattori critici del processo;
- Selezionare le metodologie di misurazione del processo;
- Attivare azioni correttive tempestive al presentarsi di deviazioni del processo;
- Sensibilizzare l'intera struttura alle tematiche del controllo di processo..

Quando un'azienda vuole certificarsi e seguire queste linee guida per la qualità diventa perciò opportuno l'impiego di tecniche statistiche per la misurazione e il controllo delle caratteristiche del processo.

Un valido strumento è il controllo statistico di processo.

Il controllo statistico di processo può essere definito come una metodologia che, in riferimento ad una determinata attività o processo caratterizzato da ripetitività, fa ricorso a tecniche statistiche al fine di definire, analizzare e verificare le condizioni che determinano la variabilità dell'oggetto di analisi.

Obiettivo di queste tecniche è quello di dotare l'impresa di strumenti adeguati per migliorare il livello dei prodotti offerti attraverso l'eliminazione di errori e difformità che causano aggiustamenti e controlli inutili e quindi rallentamenti nei cicli di lavorazione.

Il controllo statistico di processo deve avvenire attraverso l'impiego di dati statisticamente significativi e perciò analizzabili che consentano di determinare e interpretare performance e cause che determinano "cambiamenti indesiderati" rispetto al normale funzionamento del processo in analisi.

In pratica si devono elaborare dati, sui cui fondare determinate conclusioni, che

consentano di risalire alla causa prima su cui occorre intervenire per modificare le condizioni di svolgimento del processo.

Nell'individuare i fattori che generano anomalie si fa riferimento a due categorie principali di cause:

- Cause comuni o accidentali
- Cause speciali o identificabili

Queste cause introducono il concetto di variabilità che è fondamentale per la qualità.

La variabilità è un aspetto ineliminabile in ogni processo produttivo, infatti, è impossibile produrre due oggetti perfettamente identici.

E' importante, però, riuscire a ridurre il più possibile la variabilità e ad individuare le cause di anomalie indesiderate nel minor tempo possibile.

Le cause comuni o accidentali sono insite nella variabilità di un processo produttivo.

Esse insorgono casualmente durante il normale svolgimento del processo e ne determinano la fluttuazione naturale.

Le cause speciali o identificabili, invece, sono tutte quelle che determinano variabilità indesiderata rispetto al naturale svolgimento del processo.

Esse possono derivare, ad esempio, dall'uso sbagliato di un utensile o da particolari condizioni ambientali, come l'illuminazione o la temperatura.

Fin quando non si provvede ad una loro rimozione o correzione esse continueranno ad influire in maniera imprevedibile sul processo, portandolo fuori controllo.

Un processo si definisce sotto controllo statistico quando la sua variabilità è dovuta solo a cause comuni.

Una volta distinte le cause speciali da quelle comuni, si deve quindi procedere, quando possibile a rimuovere le cause speciali con lo scopo di raggiungere una condizione di stabilità.

Gli studi mirati al riconoscimento delle cause e alla prevenzione dei difetti non sono certo temi nuovi alla qualità.

Già nel 1924, infatti, il dott. W.A. Shewart iniziò a sviluppare un approccio statistico al controllo qualità osservando che il concetto di variabilità è adeguato all'analisi e alla descrizione dei processi produttivi.

Con il contributo della statistica inferenziale e descrittiva, arrivò allora alla formulazione delle sue carte di controllo, uno degli strumenti statistici più impiegati nell'analisi dei processi produttivi.

Da allora i passi avanti compiuti sul tema sono stati molti, primo fra tutti il riconoscimento della validità di questi strumenti e un loro più vasto impiego.

## **1.2 Le carte di controllo**

La prima Carta di Controllo fu proposta il 16 maggio 1924 da Walter A. Shewhart, allora capo dei Laboratori Bell, che in una nota al capo del Dipartimento di Ingegneria Ispettiva della Western Electric scriveva : *"il modello di rapporto allegato è stato progettato per indicare se le variazioni osservate nella percentuale di apparati difettosi siano o no significative, a indicare cioè se il processo sia soddisfacente"*.

Il rapporto allegato divenne in seguito noto come la prima Carta di Controllo di Shewhart.

Le carte di controllo rappresentano un valido strumento per comprendere se un processo è statisticamente sotto controllo, e, se non lo è, danno un'indicazione del motivo del fuori controllo.

Le carte di controllo hanno lo scopo di rilevare la presenza di anomalie, riuscendo così a ridurre gli sprechi ed ad aumentare la produttività.

Esse forniscono importanti informazioni sul processo produttivo e permettono di capire quando intraprendere le azioni correttive, e quando lasciare seguire al processo il suo naturale sviluppo.

Se da una parte, infatti, la carta di controllo indica il comportamento anomalo di un processo che rende necessaria un'azione correttiva, dall'altra evita controlli superflui o continue regolazioni del processo che avrebbero l'effetto indesiderato di aumentarne la variabilità.

Inoltre la carta può dare degli indizi su quale tipo di azioni correttive sono necessarie e il reale effetto che hanno avuto sul processo.

Un altro vantaggio dovuto a queste carte è che danno informazioni sulla capacità e sulla stabilità del processo che in termini di qualità diventa molto utile per le decisioni aziendali.

Infatti, quando il processo si trova in stato di controllo statistico è possibile valutarne la reale efficienza produttiva e la capacità di rientrare nelle specifiche richieste, e quindi di soddisfare il cliente.

### **1.3 Componenti delle carte di controllo**

Del processo sotto esame sono raccolti i dati necessari e da questi sono ricavati i parametri statistici come media, deviazione standard, o il range.



Essi sono poi utili per la costruzione della carta.

Sulla carta di controllo sono sempre indicati tre limiti, rappresentati da altrettante linee: il limite centrale LC, e i limiti di controllo superiore UCL ed inferiore LCL.

Essi sono determinati sulla base della variabilità naturale del processo ed hanno il compito di fornire un aiuto alla lettura della carta.

I limiti UCL e LCL rappresentano i valori critici con cui la statistica viene confrontata.

I limiti UCL e LCL sono scelti in modo tale che la probabilità che la statistica in questione cada al loro interno, sia quasi uno se il processo è sotto controllo statistico.

Se invece cade fuori dei limiti di controllo è probabile che all'interno del nostro processo sia presente una causa speciale.

La costruzione delle carte di controllo si basa sull'analisi statistica di un processo: ciò le rende soggette ad un certo grado di incertezza, del quale è sempre necessario tenere conto, e che può essere aumentato o diminuito agendo sulla posizione dei limiti di controllo sulla carta.

Questa incertezza rende possibile degli errori di interpretazione delle carte di controllo, infatti, è possibile che il processo lanci un allarme anche se il processo è sotto controllo.

Si definiscono due errori di interpretazione, gli errori di tipo I e gli errori di tipo II.

L'errore di tipo I si commette quando si suppone che un processo sia fuori controllo statistico mentre in realtà è sotto controllo.

Si consideri, ad esempio, un processo sotto controllo, cioè governato solo da un insieme di cause comuni, se la misura di un campione dovesse cadere oltre i limiti esterni, si potrebbe ipotizzare che il processo è fuori controllo.

Tuttavia, essendo i limiti ad una distanza finita dal limite centrale, esiste una probabilità, seppur piccola, che il processo sia in realtà sotto controllo.

In tale caso considerare il processo fuori controllo è sbagliato.

La probabilità di commettere un errore di tipo I è solitamente indicata con  $\alpha$ .

L'errore di tipo II, invece, si commette considerando il processo sotto controllo quando in realtà non lo è.

Ciò avviene quando nessuna osservazione cade al di fuori dei limiti di controllo, ma in realtà sono avvenuti dei cambiamenti nel processo.

Affinché la carta sia efficiente bisogna perciò ridurre la probabilità di commettere questi errori.

Quando la statistica esce dai limiti, si ha un allarme, quest'evento è il segnale di una potenziale situazione di fuori controllo.

La distinzione tra un falso allarme (errore di tipo I) ed un vero fuori controllo non è sempre riconoscibile, si cerca perciò di costruire una che segnali il minor numero di falsi allarmi possibili e che riesca a rilevare un vero fuori controllo nel minor tempo possibile.

Per questi motivi si parla di Average Run Length (ARL) che è il numero medio di campioni necessario per segnalare un allarme, quando la media del processo ha subito un salto di ampiezza  $\delta$  per lo scarto.

Per avere una carta efficiente è importante che l'ARL sia il più grande possibile quando il processo è in controllo, e che sia minima quando si ha un fuori controllo.

L'ARL dipende dalla probabilità dei due errori, dall'ampiezza  $\delta$  che si vuole rilevare e dalla numerosità campionaria che incide sulla variabilità del processo.

Altro fattore importante nel controllo di processo sono i limiti di specifica.

Dato che i limiti di controllo sono determinati sulla base della variabilità naturale del processo, per necessità esistono anche dei limiti di specifica,

indipendenti dall'andamento del processo, essi sono individuati da ingegneri, progettisti o anche dai clienti e definiscono le caratteristiche tecniche che il prodotto deve avere ai fini della produzione e della vendita.

La linea centrale dell'intervallo di specifica viene denominata livello nominale LN ed è il valore desiderato per la caratteristica in questione, ovviamente è opportuno che il livello nominale coincida con la linea centrale dei limiti di controllo.

Solitamente è appropriato che i limiti di specifica siano esterni a quelli di controllo, se ciò non si verifica si possono avere dei prodotti non conformi all'interno della regione di accettazione.

## **1.4 Carte per variabili e carte per attributi**

Le carte di controllo si dividono in due categorie in base al tipo di dati, dati variabili e dati attributi.

Si hanno dunque le carte per variabili e le carte per attributi.

I dati variabili sono misurabili su una scala numerica ( ad esempio tramite unità di lunghezza, diametro, peso, temperatura), sono continui e se riferiti alla stessa unità di misura, possono essere confrontati numericamente, permettendo di ricavare informazioni come media e dispersione.

Le carte per attributi, invece, sono basate sul conteggio dello stato delle unità e possono, ad esempio, indicare la presenza o l'assenza di una condizione, come ad esempio i difetti di lavorazione .

I dati attributi sono qualitativi e di rilevazione dicotomica.

Le carte per attributi possono essere molto utili nella rilevazione delle non conformità.

La non conformità è la non rispondenza di un certo attributo dell'oggetto prodotto alle specifiche date.

Un pezzo non conforme potrà dunque essere caratterizzato da una o più non conformità.

Il numero di dati necessari per la compilazione di una carta per attributi é solitamente minore del numero di dati necessari a una carta per variabili, queste ultime forniscono però un maggior numero di informazioni delle carte per attributi.

Nella seguente tabella sono riportati i differenti tipi di carte di controllo con riferimento al tipo di dati trattati e al parametro del processo sotto controllo.

<b>Tipo di Dati</b>	<b>Tipo di Carta</b>	<b>Parametro del processo sotto controllo</b>
Variabili	X medio	Media
	R	range medio
	S	scarto medio
	X	valori singoli
Attributi	P	frazione di non conformi
	NP	numero di non conformi
	C	numero di non conformità
	U	numero di non conformità per unità

- Le carte per variabili

Le carte per variabili sono spesso utilizzate in coppia.

Una di esse controlla la media del processo (carta  $\bar{x}$  medio), l'altra la dispersione (carta R o carta S).

La combinazione delle carte di controllo aumenta la possibilità di individuare un processo fuori controllo e, rispetto alla carta singola, fornisce una maggiore quantità di informazioni utili per eliminare le cause attribuibili.

Per questo motivo si parla di carta  $\bar{x}$  medio R o di carta  $\bar{x}$  medio S come se si trattasse di una carta sola.

Le carte  $\bar{x}$  medio e le carte R sono le più utilizzate nel controllo e nell'analisi delle variabili di un processo.

Le misurazioni inerenti ad una specifica caratteristica del processo sono raccolte in campioni di limitate dimensioni, generalmente da due a otto per ogni campione.

La costruzione di una carta  $\bar{x}$  medio R richiede il calcolo della media e del range R per ogni campione, per poterli poi rappresentare graficamente.

La sostanziale differenza con la carta S è la bassa numerosità campionaria che rende conveniente l'uso del range al posto dello scarto medio.

Anche le carte  $\bar{x}$  - medio e le carte S sono usate in congiunzione.

Lo scarto campionario  $s$  è un indicatore molto efficiente della variabilità di un processo, specialmente per campioni di grandi dimensioni (solitamente con  $n$  maggiore di otto), ma è meno facile da calcolare e meno sensibile all'anomalia di un unico valore in un campione.

Il *range*, infatti, definito come la differenza tra il valore massimo e il minimo dei valori all'interno di un campione, mette bene in evidenza valori che si allontanano eccessivamente dagli altri.

Le carte X a differenza delle carte precedenti, che raggruppano e valutano campioni di dati composti da diversi individui, sono caratterizzate dall'analisi di misure singole.

Ogni campione è composto di un unico elemento, situazione tipica delle procedure di raccolta dati molto costose o quando il tempo di fabbricazione è molto lungo.

Per una carta X l'indice di dispersione viene calcolato utilizzando la variazione mobile, definita come il valore assoluto della differenza tra due osservazioni successive.

Non calcolando medie, infatti, il teorema del limite centrale non è applicabile, a differenza delle carte precedenti.

- Le carte per attributi

#### 1. Carta P e carta nP

La carta p è utilizzata per mostrare la frazione di elementi non conformi di un campione di grandezza costante.

La Carta np, a differenza della carta p, misura il numero di unità non conformi nel campione, anziché la frequenza.

#### 2. Carta C e carta U

La carta c è quella principalmente utilizzata nell'analisi dei dati di tipo attributo. Questa carta, a differenza delle carte P e nP, misura il numero totale di difetti per campione, anziché il numero di elementi difettosi.

Un elemento del campione, infatti, può avere più di un difetto, perciò questa carta permette di controllare più attributi contemporaneamente.

La carta c deve avere numerosità campionaria costante.

La carta u, invece, è l'unica di queste utilizzata per campioni formati da un numero variabile di elementi.

Essa misura il numero medio di difetti per unità di riferimento, come ad esempio il numero di difetti per cm<sup>2</sup> di stoffa.

## 1.5 La carta Shewhart

La carta di controllo di tipo Shewhart è una carta senza memoria, cioè una carta che tiene conto solo delle informazioni correnti e non di quelle passate.

Questa carta è composta dalla statistica di controllo  $w$  che misura la caratteristica da monitorare e dai limiti di controllo.

La statistica di controllo  $w$  è data da:

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

I limiti di controllo sono invece tre, rispettivamente superiore, centrale e inferiore, e soddisfano le seguenti relazioni:

$$UCL = \mu_w + L\sigma_w$$

$$LC = \mu_w$$

$$LCL = \mu_w - L\sigma_w$$

In queste formule  $w$  rappresenta la statistica di controllo, la misura che si intende monitorare,  $\mu_w$  rappresenta la sua media e  $\sigma_w$  lo scarto del processo.

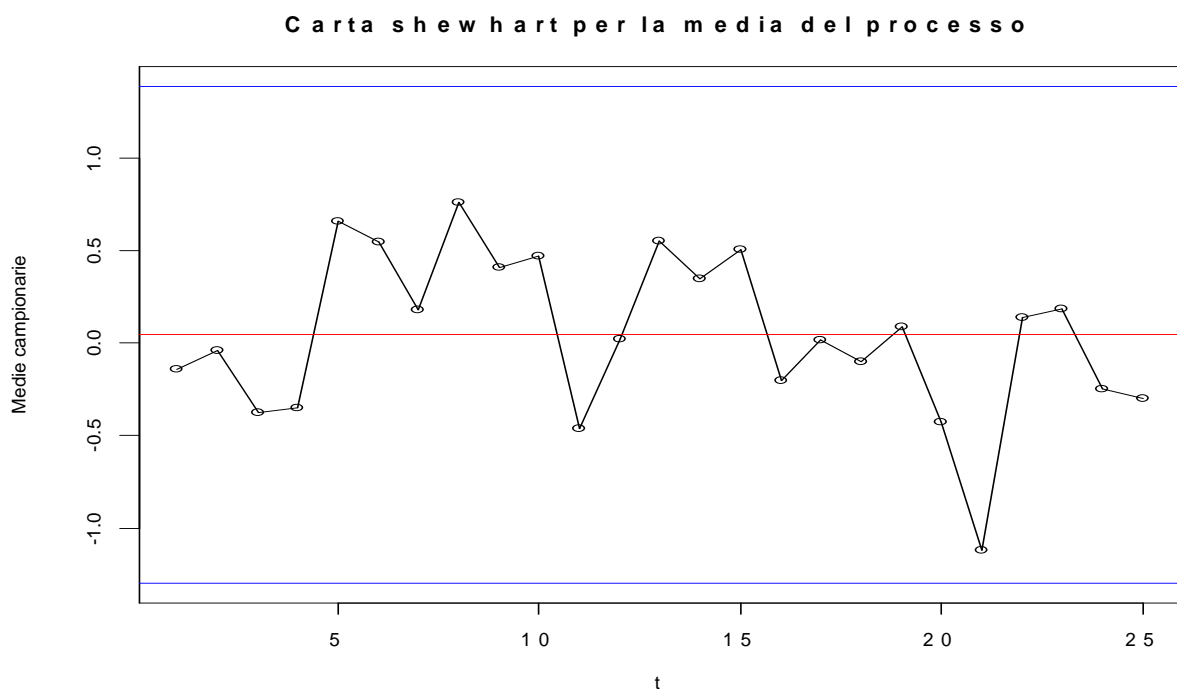
$L$ , invece, è una costante che moltiplicata per lo scarto fornisce la distanza dei limiti di controllo LCL e UCL dalla linea centrale LC che generalmente è la media della statistica campionaria  $w$ .

La media della statistica di controllo ( $\mu_w$ ) deve essere uguale ad un valore considerato ottimale, in altre parole deve essere conforme alle specifiche.

Il seguente grafico è una carta Shewhart per la media del processo, nella quale ho inserito 125 dati presi da una normale con media 0 e varianza 1, divisi in campioni di 5 elementi.

Si hanno così 25 valori della statistica campionaria.

Ho posto il parametro L uguale a tre, scelta molto diffusa che permette un'ARL in controllo di 370.



Un limite di questa carta è che può essere utilizzata solo nel caso che i dati rispettino un'assunzione di normalità.

## 1.6 La carta EWMA

La carta di controllo EWMA (Exponentially Weighted Moving Average) è una carta con memoria, cioè usa le informazioni ottenute dai dati precedenti.



Questo tipo di carta è usata per notare eventuali cambiamenti nella media del processo ed in modo specifico nei confronti di piccoli salti di livello per i quali le carte Shewhart sono poco affidabili (solitamente per salti minori di  $1.5\sigma$ ). In compenso la carta EWMA non reagisce velocemente come la carta Shewhart per i salti di livello elevato.

Il peso delle informazioni passate ne riduce, infatti, la reattività, ma nel caso di un salto nella media, nel grafico si comincia a notare un trend crescente o decrescente, secondo la variazione.

Inoltre a differenza della Shewhart la carta EWMA viene solitamente adoperata per le osservazioni singole, l'uso di campioni o sottogruppi ne riduce l'efficienza.

Può quindi risultare utile combinare l'utilizzo della carta di Shewhart con quella EWMA.

La carta EWMA, ideata da Roberts nel 1959, in poche parole è una media ponderata, a pesi decrescenti secondo una progressione geometrica, di tutte le osservazioni passate e di quella corrente.

La statistica di controllo per la carta EWMA a misure singole è la seguente:

$$w_t = (1 - \lambda)w_{t-1} + \lambda(x_t)$$

Mentre per la carta EWMA con numerosità campionaria maggiore di uno, la statistica di controllo si calcola:

$$w_t = (1 - \lambda)w_{t-1} + \lambda g(\underline{x}_t) \quad 0 < \lambda \leq 1$$

In queste formule si nota un parametro nuovo,  $\lambda$ , detto costante di lisciamiento, questo serve per pesare le informazioni dei dati passati rispetto a quelli correnti.

Si può notare, infatti, che se si pone  $\lambda$  uguale a uno si ha la statistica di controllo di una carta Shewhart:

$$w_t = 0 \cdot w_{t-1} + 1 \cdot g(\underline{x}_t) \quad \rightarrow \quad w_t = g(\underline{x}_t)$$

Con  $\lambda$  vicini allo zero, invece, i dati passati hanno una notevole influenza sulla statistica di controllo.

Se  $0 < \lambda < 1$ , allora la carta EWMA prende in considerazione anche le informazioni passate  $w_{t-1}$ .

Se sostituiamo ripetutamente i  $w_{t-1}$  con i loro valori si ottiene:

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

I pesi  $(1-\lambda)^i$  decrescono in progressione geometrica e la loro somma è sempre pari a all'unità.

La varianza della statistica EWMA, nel caso i dati siano determinazioni di variabili casuali indipendenti di varianza comune  $\sigma^2$ , si trova grazie a:

$$\sigma_{w_i}^2 = \sigma^2 \left( \frac{\lambda}{2 - \lambda} \right) \left[ 1 - (1 - \lambda)^{2i} \right]$$

Da questa formula si può notare che (dati  $\sigma^2$  e  $\lambda$  come costanti)  $\sigma_{w_i}^2$  cresce in maniera asintotica all'aumentare degli istanti  $i$ .

Per  $i \rightarrow \infty$  il termine  $[1 - (1-\lambda)^{2i}]$  dà 1, perciò il valore asintotico della varianza risulta:

$$\sigma_{w_i}^2 = \sigma^2 \left( \frac{\lambda}{2 - \lambda} \right)$$

I limiti di controllo per la carta EWMA sono:

$$UCL = \mu_w + L \sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right)} \left[ 1 - (1 - \lambda)^{2i} \right]$$

$$LC = \mu_w$$

$$LCL = \mu_w - L \sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right)} \left[ 1 - (1 - \lambda)^{2i} \right]$$

Si può notare che i limiti non sono costanti come nella carta Shewhart, ma sono asintotici perché dipendono dallo scarto  $\sigma_{w_i}$ .

I valori asintotici dei limiti sono:

$$UCL = \mu_w + L \sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right)}$$

$$LC = \mu_w$$

$$LCL = \mu_w - L \sigma \sqrt{\left( \frac{\lambda}{2 - \lambda} \right)}$$

Ma per disegnare i limiti bisogna individuare la combinazione ottimale per i parametri L e  $\lambda$ .

Dato un ARL in controllo, i parametri L e  $\lambda$  sono determinati in modo che l'ARL fuori controllo sia il più piccolo possibile per un salto di grandezza fissata.

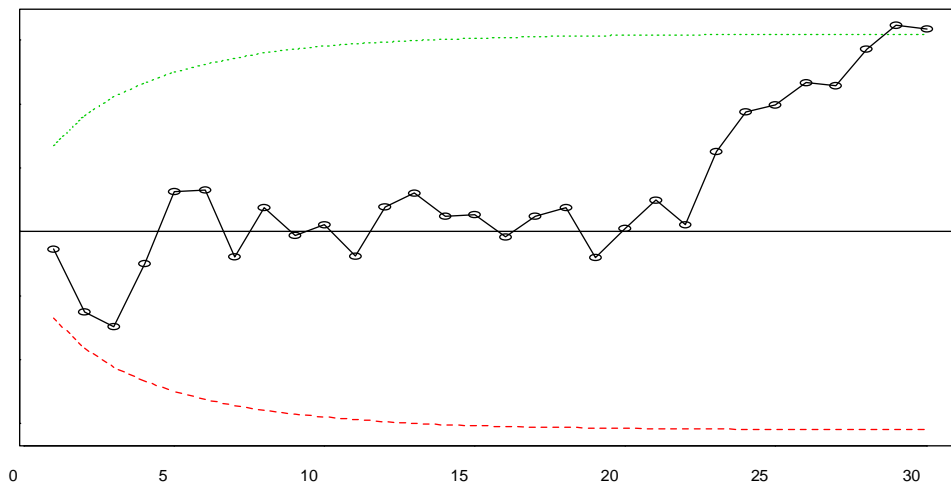
Sono stati effettuati numerosi studi sulla scelta di questi parametri, Crowder prima e Lucas e Saccucci poi, hanno fornito delle tabelle dalle quali si ricavano valori di ARL dati i due parametri.

Generalmente si può affermare che si hanno buone prestazioni della carta EWMA nel rilevare piccoli salti di livello con  $\lambda$  compresi tra 0.05 e 0.25.

Dal grafico si può notare che quando il processo è in controllo la statistica oscilla attorno al valor medio, mentre quando avviene un salto di livello si evidenzia un trend che porta i valori a uscire dai limiti.

L'individuazione di cambiamenti nel processo, perciò, può avvenire rilevando la presenza di andamenti crescenti o decrescenti della statistica di controllo, anche se quest'ultima non è ancora uscita dai limiti.

### Esempio di una carta EWMA con fuori controllo



Come si può vedere dal grafico, la carta segnala la presenza di un fuori controllo ad un istante preciso (in questo caso nell'istante 29).

A questo punto è utile procedere con la stima dell'istante in cui è avvenuto il cambiamento.

Generalmente tale istante è definito come il punto in cui la statistica EWMA incrocia la media del processo

Un altro vantaggio che la carta EWMA possiede nei confronti della Shewhart, è che essa non deve per forza rispettare l'assunzione di normalità dei dati.

La carta EWMA può essere utilizzata direttamente sui dati tra loro correlati, che sono molto diffusi nella realtà aziendale.

## 1.7 Le carte di controllo non parametriche

Solitamente si suppone che i dati siano distribuiti in modo parametrico, o più comunemente in modo normale.

Alcune carte di controllo, come le Shewhart, risultano corrette solo se è soddisfatta l'assunzione di normalità.

L'impiego della normalità è particolarmente ragionevole se il processo di controllo è basato su campioni di almeno quattro o cinque osservazioni.

Lo stesso teorema del limite centrale afferma che la distribuzione della somma di un numero elevato di variabili casuali indipendenti e identicamente distribuite tende a distribuirsi normalmente, indipendentemente dalla distribuzione delle singole variabili.

A volte però ci sono ragioni tecniche ed economiche che costringono l'uso di osservazioni singole, ed in questo caso il teorema del limite centrale non può più venire in aiuto.

Per superare questo problema sono state disegnate le carte di controllo non parametriche che si basano sulla distribuzione della run length in controllo.

Se un processo è in controllo, la distribuzione della run length è la stessa per qualsiasi distribuzione continua.

Le carte non parametriche sono meno efficienti rispetto alle corrispondenti carte parametriche, ma permettono una maggiore elasticità dato che non c'è bisogno di ipotizzare alcuna distribuzione per la statistica da controllare.

Le carte di controllo non parametriche, anche se meno precise, permettono di conseguire numerosi vantaggi come:

- Semplicità, sono facili da costruire
- Non c'è bisogno di individuare una particolare distribuzione per la statistica di controllo

- Distribuzione dell'ARL in controllo uguale per tutte le distribuzioni continue
- Robustezza nei confronti dei valori anomali

## CAPITOLO 2

### CARTA EWMA BASATA SUI RANGHI

#### 2.1 Introduzione

Nel 1991 Hackl e Ledolter hanno proposto delle carte di controllo non parametriche basate sui ranghi.

Il rango standardizzato  $R_t$  si definisce con la seguente formula:

$$R_t = 2[F_0(x_t) - 1/2]$$

Dove  $X_t$  è una sequenza di variabili casuali indipendenti da una distribuzione di probabilità continua  $F(x)$ .

Il processo è considerato in controllo quando  $F(x) = F_0(x)$

$R_t$  ha distribuzione uniforme  $[-1;1]$  con media 0 e varianza  $1/3$ .

I ranghi sono indipendenti.

Nel caso la distribuzione in controllo non sia nota si può usare un campione storico di ampiezza  $g-1$ ,  $\{Y_1, Y_2, \dots, Y_{g-1}\}$ , preso da un processo in controllo e definire il rango standardizzato di  $X_t$  come:

$$R_t = \frac{2}{g} \left[ R_t^* - \frac{g+1}{2} \right]$$

dove  $R_t^*$  è il rango di  $X_t$  rispetto il campione  $\{Y_1, Y_2, \dots, Y_{g-1}\}$ .

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

dove la funzione  $I$  è uguale a 1 se  $X_t > Y_i$  e 0 altrimenti

Il rango  $R_t$  segue una distribuzione discreta uniforme sui  $g$  punti  $\{(1/g)-1, (3/g)-1, \dots, 1-(1/g)\}$  con media 0 e varianza  $(g^2-1)/3g^2$ .

Si può notare che usando un campione di riferimento viene meno la proprietà di indipendenza dei ranghi, infatti usando la funzione I si può trovare che la correlazione  $\text{Corr}(R_t, R_{t+k}) = (g+1)^{-1}$  per tutti i k.

All'aumentare di g, comunque, questa correlazione tende a 0.

Hackl e Ledolter propongono una carta basata su una statistica EWMA dei ranghi.

Per ogni periodo t si calcola la seguente statistica di controllo:

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

Il valore iniziale  $T_0$  è solitamente considerato uguale a 0 e  $\lambda$  è il parametro di lisciamiento.

La carta EWMA basata sui ranghi dipende fondamentalmente da due parametri: il parametro di lisciamiento  $\lambda$  ed il limite di controllo h.

Questi due parametri devono essere scelti in modo che l'ARL sia grande se il processo è in controllo e piccolo nel caso di fuori controllo.

## 2.2 ARL per processi in controllo

Per ovviare a questo problema, Hackl e Ledolter, basandosi sull'equazione di Crowder (applicata ai ranghi), sono riusciti a ricavare una tabella che per ogni coppia di parametri,  $\lambda$  e h, fornisce un'ARL.

Confrontando le ARL ottenute con quelle di un'EWMA e di una Shewhart basate sulle osservazioni, si nota che le prestazioni della carta EWMA basata sui ranghi sono certamente migliori.

Per avvalorare questa impressione Hackl e Ledolter hanno proseguito il loro studio con ulteriori simulazioni e sono riusciti ad ottenere le stime di Monte



Carlo di un' ARL con  $\lambda=0.3$  e  $h=0.50$  con campioni di riferimento di ampiezza  $g=30$ ,  $g=50$  e  $g=100$ . (All'aumentare di  $g$  si riduce la correlazione)

I risultati indicano che se si stimasse la distribuzione l' ARL potrebbe essere sostanzialmente più grande.

Con questa combinazione di parametri, quando la procedura è basata sui ranghi come definiti inizialmente  $R_t = 2[F_0(x_t) - 1/2]$  si ha un' ARL di 102.

Per le carte basate sulle osservazioni il campione di riferimento è utile per il calcolo della deviazione standard  $s_g$ , che serve anche per il calcolo del limite di controllo  $h=c*s_g$  della carta EWMA basata sui ranghi.

Il moltiplicatore  $c$  viene scelto in modo che l' ARL sia 102 quando viene usata la deviazione standard nota.

Per paragonare le prestazioni delle varie carte la distribuzione della variabile di controllo gioca però un ruolo determinante.

Hackl e Ledolter hanno perciò proseguito il loro studio testando la loro carta su osservazioni generate da diverse distribuzioni come una  $t(4)$  e una uniforme.

In particolare risulta che la carta EWMA funziona bene per valori di  $\lambda$  compresi tra 0.1 e 0.3.

### **2.3 ARL per processi fuori controllo**

Per comprendere l'utilità della carta EWMA basata sui ranghi è necessario studiarne l'efficacia nel rilevare cambiamenti nel processo.

Hackl e Ledolter continuano perciò il loro studio prestando attenzione alle diverse distribuzioni, a differenti campioni di riferimento ed attuando numerose simulazioni, in un procedimento simile al precedente.

I risultati mostrano che, in termini di ARL, c'è una piccola perdita di efficienza per variabili di controllo distribuite normalmente o uniformemente, ma che questa è notevolmente compensata dai vantaggi ottenuti nel caso di distribuzioni a code pesanti soprattutto se i salti di livello sono piccoli o moderati.

## **2.4 Conclusioni**

I risultati dimostrano che, nel caso di distribuzioni normali o uniformi, c'è un basso costo da pagare nell'usare la carta basata sui ranghi rispetto a quelle basate sulle osservazioni.

A fronte di questo però la carta EWMA basata sui ranghi permette delle buone prestazioni nel caso di piccoli o moderati salti di livello, e nel caso di distribuzioni non normali.

La carta Ewma applicata ai ranghi standardizzati a differenza delle carte classiche è meno sensibile agli outliers (cioè a singole osservazioni estreme) ed allo stesso tempo perciò è meno efficace nel rilevare velocemente ampi salti di livello.

## **CAPITOLO 3**

# **CARTA EWMA PER MONITORARE LA MEDIA DI UN PROCESSO CARATTERIZZATO DA DISTRIBUZIONE DI WEIBULL**

### **3.1 Introduzione**

A volte i dati non normali si possono adattare molto bene ad altre distribuzioni, questo è il caso della distribuzione di Weibull, brevemente sintetizzata qui di seguito, che tratta un particolare gruppo di dati.

### **3.2 La distribuzione di Weibull**

La distribuzione di Weibull nasce in ingegneria come modello teorico della distribuzione dei tempi di durata di vari dispositivi elettrici o elettronici, quali un diodo o anche una semplice lampadina, che restano attivi per un tempo aleatorio  $T$  dopo il quale smettono di funzionare.

La distribuzione di Weibull è diffusa perché si adatta in modo specifico alle popolazioni di dati che hanno come caratteristica la durata.

Per comprendere meglio questa particolare distribuzione è meglio parlare di Weibull semplice e Weibull generalizzata.

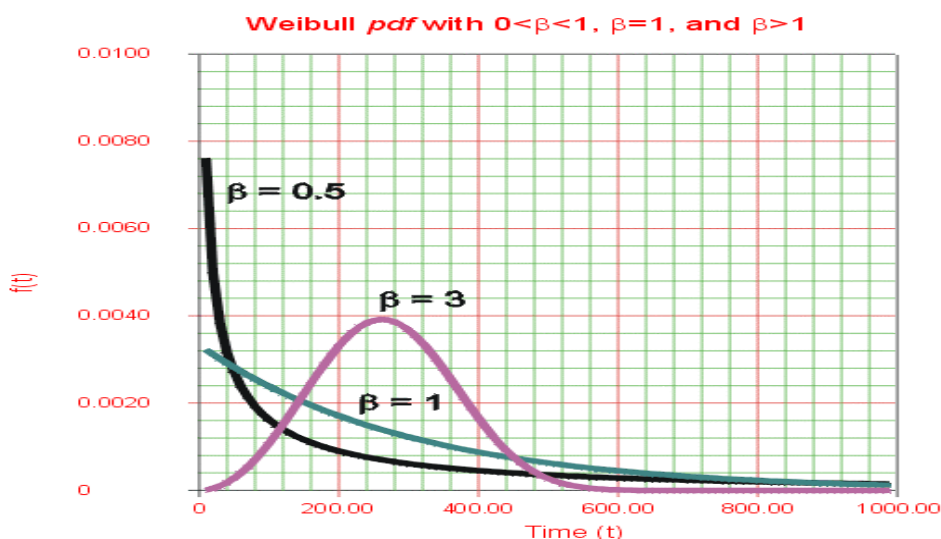
- La distribuzione di Weibull semplice

La funzione di densità della distribuzione Weibull semplice presenta un unico parametro ed è la seguente:

$$f(t) = \beta t^{\beta-1} \exp(-t^\beta) \quad t > 0, \beta > 0$$

Il parametro  $\beta$  viene chiamato parametro di forma, infatti al mutare del suo valore la funzione assume differente configurazione:

- se  $0 < \beta < 1$ . f è a forma di U
- se  $\beta = 1$ . f è decrescente
- se  $\beta > 1$  con moda a  $[(\beta - 1) / \beta]^{1/\beta}$  f è unimodale



La sua funzione di ripartizione è data da  $F(t) = P(X \leq t)$  per  $x$  appartenente a  $\mathbf{R}$  ed è:

$$F(t) = 1 - \exp(-t^\beta) \quad t > 0.$$

La sua funzione quantile è

$$F^{-1}(p) = [-\ln(1 - p)]^{1/\beta} \quad 0 < p < 1.$$

La sua funzione di affidabilità è

$$G(t) = \exp(-t^\beta) \quad t > 0.$$

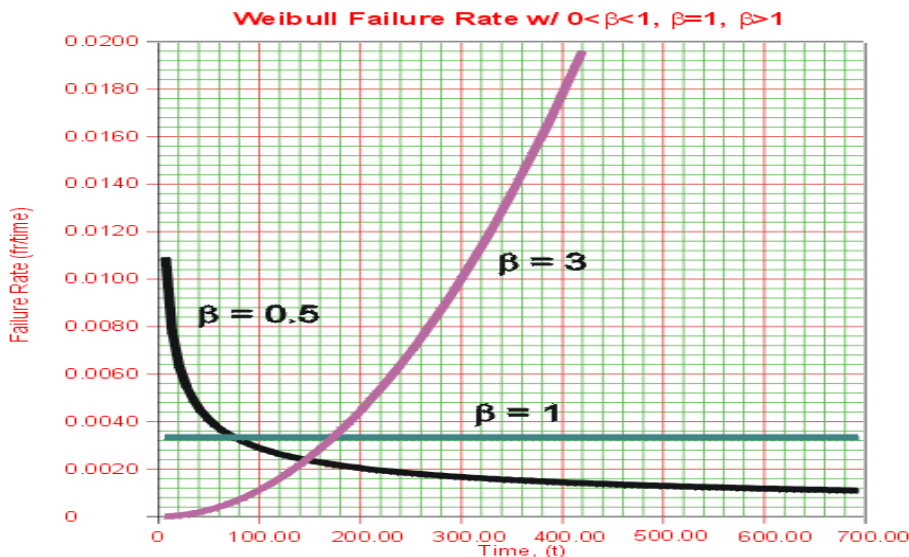
La funzione tasso di guasto (censoring rate) è

$$h(t) = \beta t^{\beta-1} \quad t > 0.$$

La funzione tasso di guasto rappresenta i dati eliminati in ogni campione.

La distribuzione di Weibull può essere applicata a congegni con tasso di guasto crescente, costante o decrescente. Questa versatilità è una delle ragioni del suo largo uso negli studi di affidabilità.

In questo grafico possiamo notare il comportamento del tasso di guasto nel tempo al variare di  $\beta$ .



- La distribuzione di Weibull generalizzata

Si usa generalizzare la distribuzione di Weibull introducendo un parametro di scala  $\alpha$ .

Così, se  $Z$  ha distribuzione di Weibull con parametro di forma  $\beta$ , allora  $X = \alpha Z$  ha distribuzione di Weibull con parametro di forma  $\beta$  e parametro di scala  $\alpha$ .

La distribuzione di Weibull  $W(\alpha, \beta)$  con parametro di forma  $\beta > 0$  e parametro di scala  $\alpha > 0$  ha densità con dominio  $\mathbb{R}^+$  data dalla formula

$$f(t) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\beta \right]$$

Si può osservare che se  $\beta = 1$ , la distribuzione di Weibull si riduce a una distribuzione esponenziale con parametro di scala  $\alpha$ .

Nel caso in cui  $\beta = 2$  si parla di distribuzione di Rayleigh con parametro di scala  $\alpha$ .

La funzione cumulativa di probabilità (ripartizione), sempre con dominio  $\mathbb{R}^+$  è descritta dalla formula:

$$F(t) = 1 - \exp \left[ - \left( \frac{t}{\alpha} \right)^\beta \right]$$

Il valor medio e deviazione standard sono invece:

$$\mu = \mu(\alpha, \beta) = \alpha \Gamma \left( 1 + \frac{1}{\beta} \right)$$

$$\sigma = \sigma(\alpha, \beta) = \alpha \sqrt{\Gamma \left( 1 + \frac{2}{\beta} \right) - \Gamma^2 \left( 1 + \frac{1}{\beta} \right)}$$

Essi sono espressi esplicitamente in termini della funzione  $\Gamma$  di Eulero:

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

Spesso accade che per un esperimento sia stanziato un budget piuttosto limitato, il quale costringe l'azienda alla scelta di limitare i costi relativi ai test.

Le decisioni nell'ambito del controllo qualità sono orientate, sulla base di valutazioni economiche, alla durata del test e al numero di unità da implementare nell'analisi sperimentale.

A tal proposito, una possibile scelta che lo sperimentatore può prendere in considerazione per contenere i costi di un esperimento è la censura della sperimentazione.

La scelta della censura della sperimentazione risulta essere per lo sperimentatore una soluzione ad un problema economico che si potrebbe avere nell'indagine di un campione completo in cui si osservano tutti i tempi al guasto degli elementi campionari.

La censura di un test può non essere esclusivamente dettata da aspetti economici.

Lo sperimentatore può decidere di rimuovere alcune unità dall'esperimento prima del guasto per poter utilizzare queste unità come campioni-test per un altro esperimento; possono inoltre influire sulla censura di un test aspetti puramente temporali, ossia la necessità o la decisione di un compromesso tra il desiderio di poter includere nel test un numero sufficiente di valori campionari e la domanda di un test rapido in termini di tempo.

### **3.3 Carta EWMA per monitorare la media di un processo caratterizzato da distribuzione di Weibull**

In molti studi medici ed industriali, a causa di limitazioni di tempo o di costo, le osservazioni vengono spesso censurate.

Nel loro studio, Zhang e Chen, riprendono alcune applicazioni di Steiner e MacKay mirate a monitorare processi industriali nei quali la censura cade ad un livello fisso.

Un esempio, usato anche da Steiner e MacKay, può essere un test sulla capacità antiruggine di un colore. Dei pannelli dipinti con un colore vengono inseriti in una camera ad una temperatura costante e vengono lasciati lì fino a che non compare la ruggine o fino ad un massimo di venti giorni.

In questo esempio le osservazioni dette censurate, sono quelle che non arrivano ad una durata di venti giorni.

Zhang e Chen usano T per indicare la durata della variabile di interesse e C per indicare il livello fisso di censura.

Supponendo che la variabile T possa essere distribuita come una Weibull, la sua funzione di densità risulta:

$$f(t, \alpha, \beta) = \frac{\beta}{\alpha^\beta} t^{\beta-1} \exp \left[ - \left( \frac{t}{\alpha} \right)^\beta \right] \quad \text{con } t > 0$$

dove  $\alpha$  è il parametro di scala e  $\beta$  il parametro di forma.

La media di T è:

$$\mu = \mu(\alpha, \beta) = \alpha \Gamma \left( 1 + \frac{1}{\beta} \right)$$

La probabilità di censura è:

$$p_c = \exp \left[ - \left( \frac{t}{\alpha} \right)^\beta \right]$$

Definendo una nuova variabile x come

$$x = \left( \frac{T}{\alpha} \right)^\beta$$

si ha che questa è distribuita come una esponenziale standard di media uguale ad uno.

Steiner e MacKay hanno dimostrato che le procedure di monitoraggio tradizionali non sono appropriate per i dati di durata, specialmente nel caso la



probabilità di censura sia elevata, ed hanno studiato delle carte di controllo Shewhart basate su valori condizionali.

Con i dati di durata è necessario prendere atto del diverso significato che assumono i cambiamenti della media. Ad un decremento della media corrisponde una diminuzione di durata e perciò di qualità, ma ad un incremento della media trova riscontro un aumento di durata e di conseguenza di qualità. E' comunque importante rilevare aumenti nella media del processo, perché questi possono suggerire nuove modifiche o nuove tecniche per apportare miglioramenti.

Lo studio di Zhang e Chen, per monitorare questo tipo di processi, propone una carta EWMA basata sui valori condizionali.

Per semplicità si assume che il tasso di censura ed i parametri  $\alpha$  e  $\beta$  della distribuzione Weibull siano conosciuti.

Si suppone, inoltre, che le osservazioni non censurate di  $T$  siano utilizzabili per comporre dei sottogruppi indipendenti di ampiezza  $n$  chiamati  $T_{i1}, T_{i2}, \dots, T_{in}$ , con  $i/1$ .

I valori attesi condizionali di  $x$  sono definiti come:

$$CE(X) = E\left(X \mid X > \left(\frac{C}{\alpha_0}\right)^{\beta_0}\right) = \left(\frac{C}{\alpha_0}\right)^{\beta_0} + 1 = -\ln(p_c) + 1$$

Rispettivamente definiamo la scala CEV

$$W_{ij} = \begin{cases} \left(\frac{T_{ij}}{\alpha_0}\right)^{\beta_0} & \text{se } T_{ij} \leq C \text{ non censurate} \\ CE(X) & \text{se } T_{ij} > C \text{ censurate} \end{cases}$$

La carta EWMA viene suddivisa in due lati, parte alta (per gli aumenti della media) e parte bassa (per i decrementi della media).

La statistica della parte bassa della carta EWMA CEV è:

$$\begin{cases} Q_i^L = \min \{ (1 - \lambda)Q_{i-1}^L + \lambda \overline{W}_i, w_0 \} \\ Q_0^L = w_0 \end{cases}$$

dove

$$\overline{W}_i = \frac{W_{i1} + W_{i2} + \dots + W_{in}}{n}$$

ed il parametro di liscio  $\lambda$  deve essere compreso tra 0 e 1 e  $w_0 = E(X) = 1$

Il limite di controllo di questa carta è:

$$LCL = K_L w_0$$

Dove  $K_L \in (0, 1)$  è un moltiplicatore che, assieme con  $\lambda$ , controlla la performance della carta.

La statistica della parte alta della carta EWMA CEV analogamente è:

$$\begin{cases} Q_i^U = \max \{ (1 - \lambda)Q_{i-1}^U + \lambda \overline{W}_i, w_0 \} \\ Q_0^U = w_0 \end{cases}$$

Ed il limite di controllo è:

$$UCL = K_U w_0$$

Dove  $K_U > 1$  è un moltiplicatore analogo a  $K_L$

## CAPITOLO 4

### CONFRONTO TRA LE DUE CARTE EWMA

#### 4.1 Introduzione

Per testare i comportamenti delle due carte EWMA le ho costruite grazie al programma R e provate su tre tipi di dati con distribuzioni diverse.

Le ho confrontate, inoltre, con una carta EWMA tradizionale.

Ho posto  $\lambda = 0.3$  in tutte le prove.

#### 4.2 Le carte EWMA con dati caratterizzati da distribuzione di

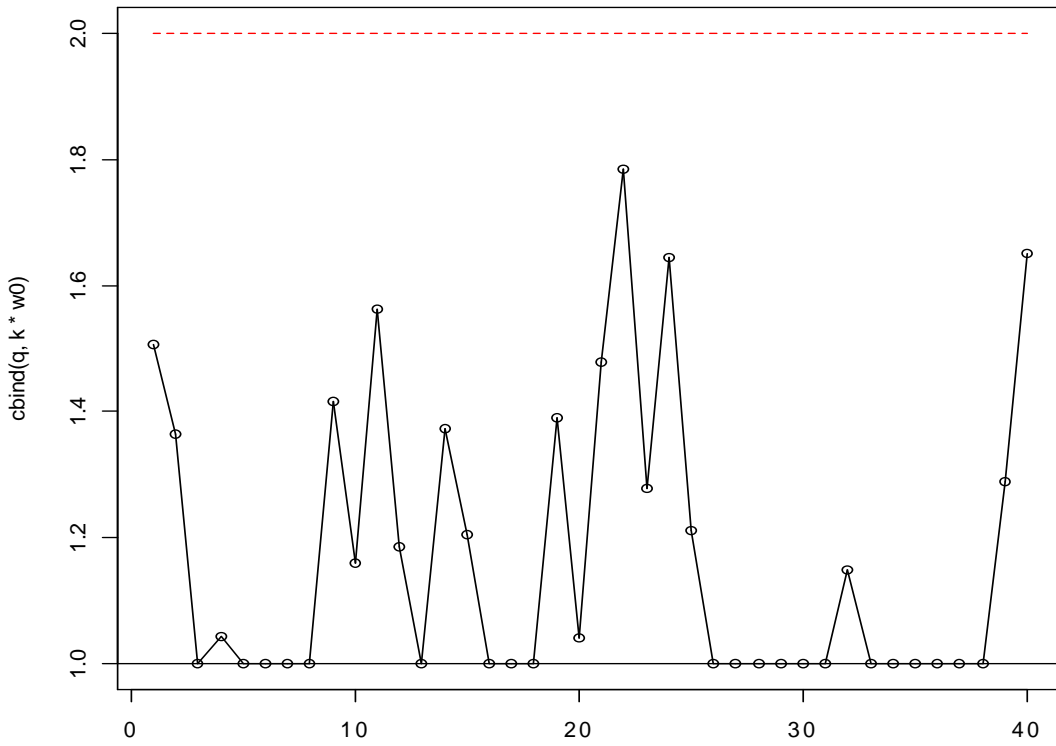
##### Weibull

Grazie al programma R ho creato 40 osservazioni prese da una Weibull con parametro di forma  $\beta=1$  e parametro di scala  $\alpha=1$ .

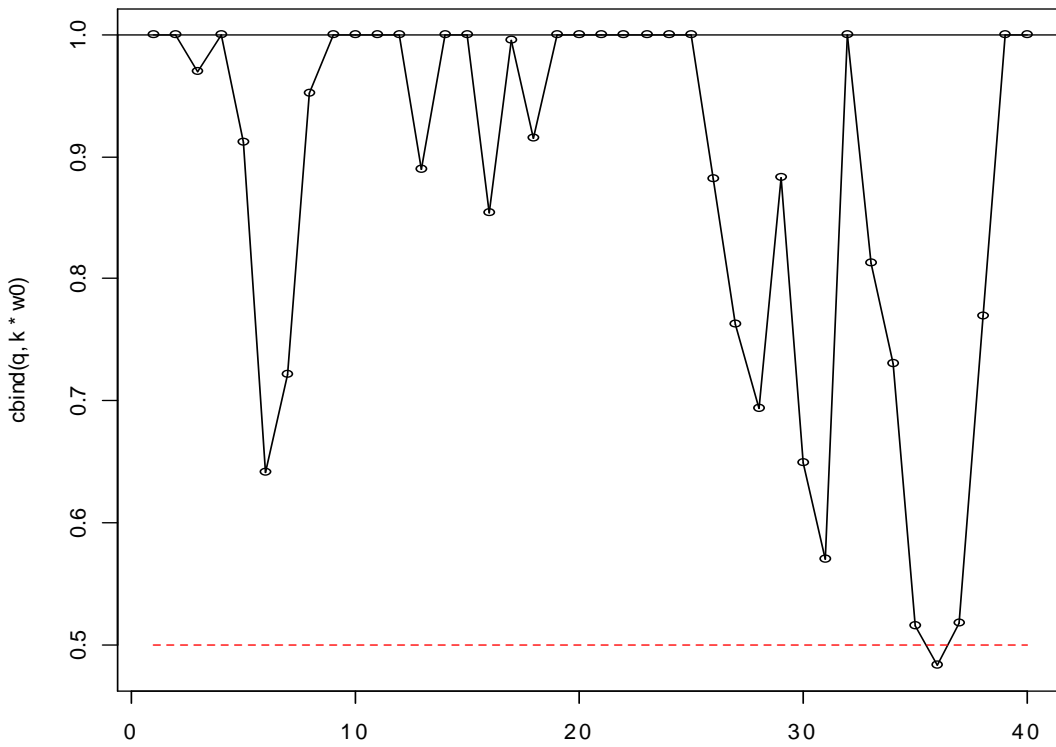
Dopo di che ho posto come livello fisso  $C=1.5$  che censura il 77,7% delle osservazioni.

Ho trasformato i dati come descritto da Zhang e Chen, ho posto  $K_L=0.5$  e  $K_U=2$  ed ho ottenuto le seguenti carte EWMA:

**carta EWMA per distribuzione Weibull parte superiore**

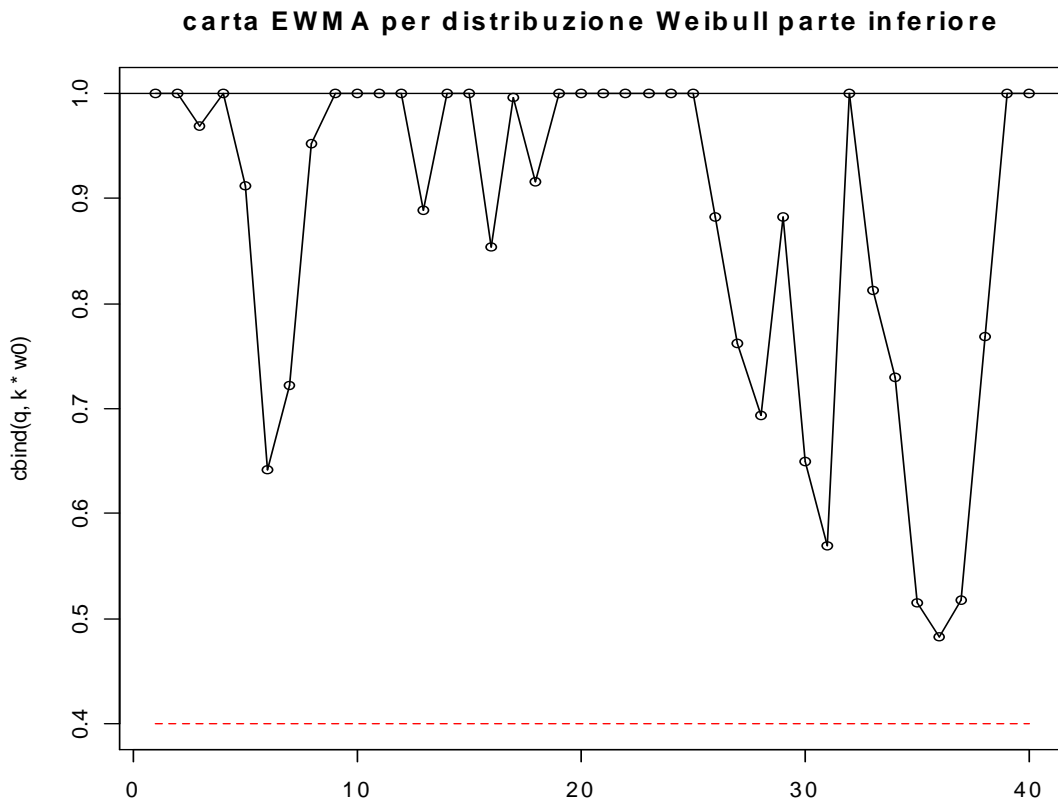


**carta EWMA per distribuzione Weibull parte inferiore**

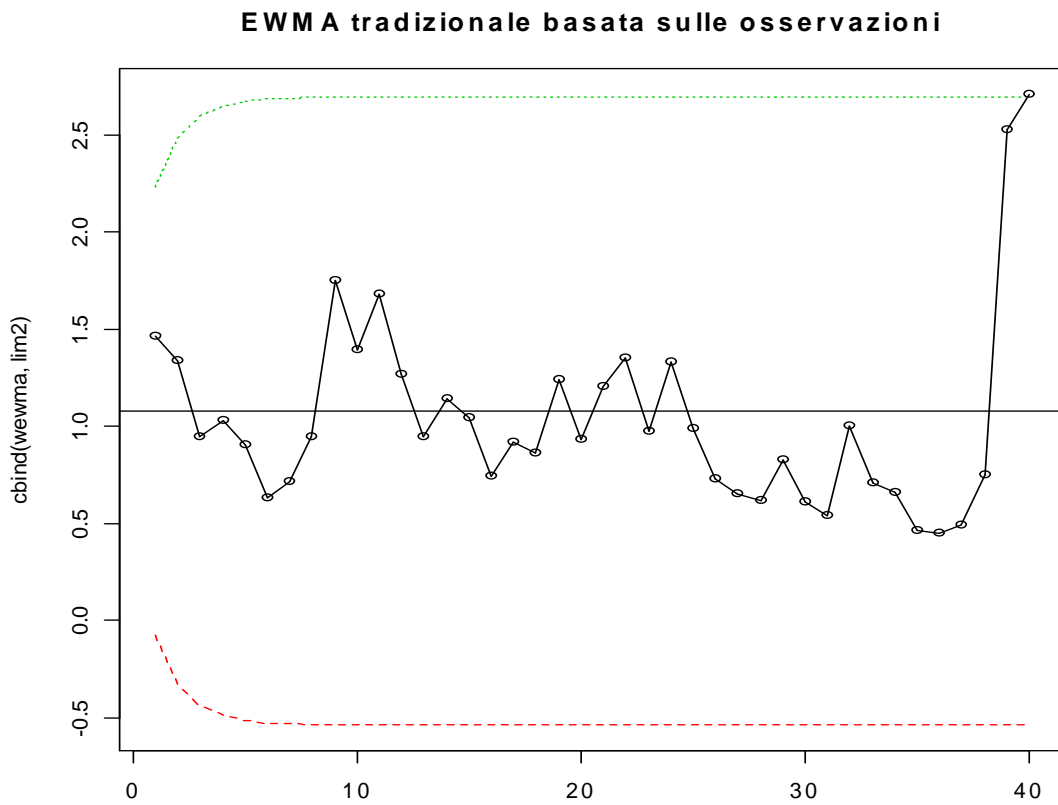


Nella parte bassa abbiamo un fuori controllo dovuto anche alla scelta di  $K_L=0.5$  che comporta una percentuale di errore del 7.5% circa, contro la scelta di  $K_U=2$  che implica un errore del 0.5% circa.

Con un  $K_L=0.4$  che comporta una percentuale di errore di circa 2.9%, la carta risulta essere già in controllo.

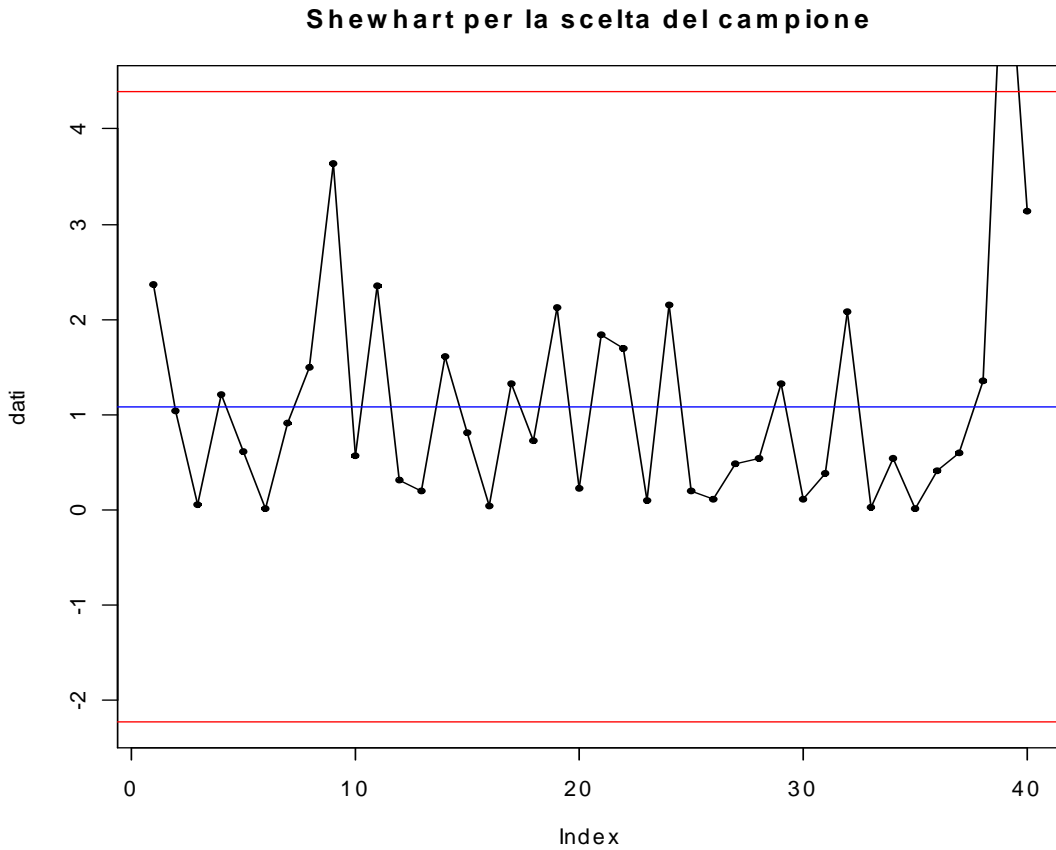


Sugli stessi dati, non trasformati, ho applicato la carta EWMA tradizionale:

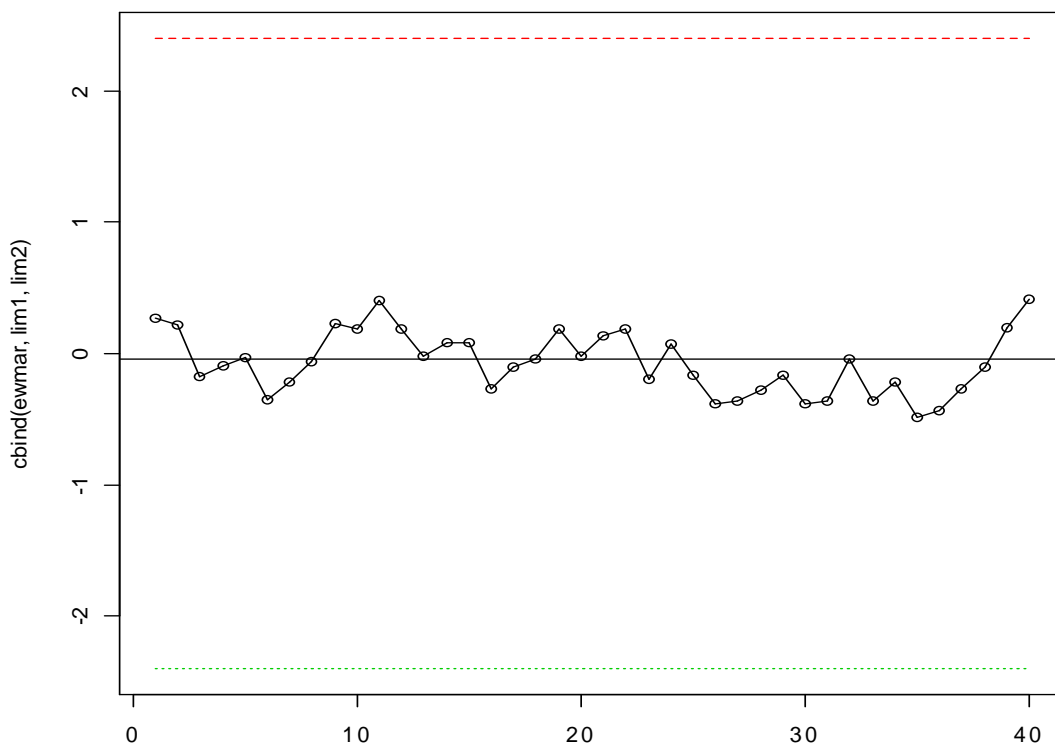


La carta EWMA tradizionale rileva già un errore alla quarantesima osservazione.

Per la carta EWMA basata sui ranghi ho utilizzato i soliti dati, ed ho applicato una Shewhart per scegliere un campione in controllo con cui confrontare i dati. Infine ho preso un campione di numerosità 10 comprendente le osservazioni dalla 20 alla 29, ed ho applicato la carta di Hackl e Ledolter.



### EWMA basata sui ranghi



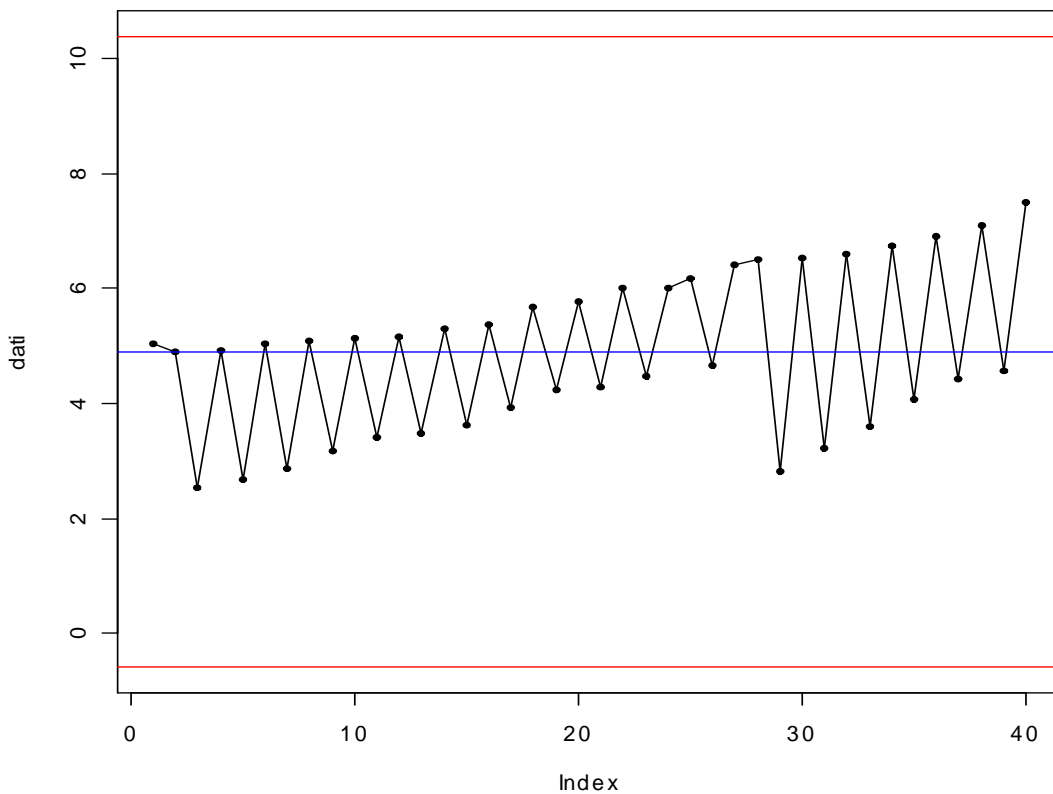
L'EWMA basata sui ranghi sembra essere molto robusta anche nei confronti della distribuzione di Weibull.

### 4.3 Le carte EWMA con dati non normali

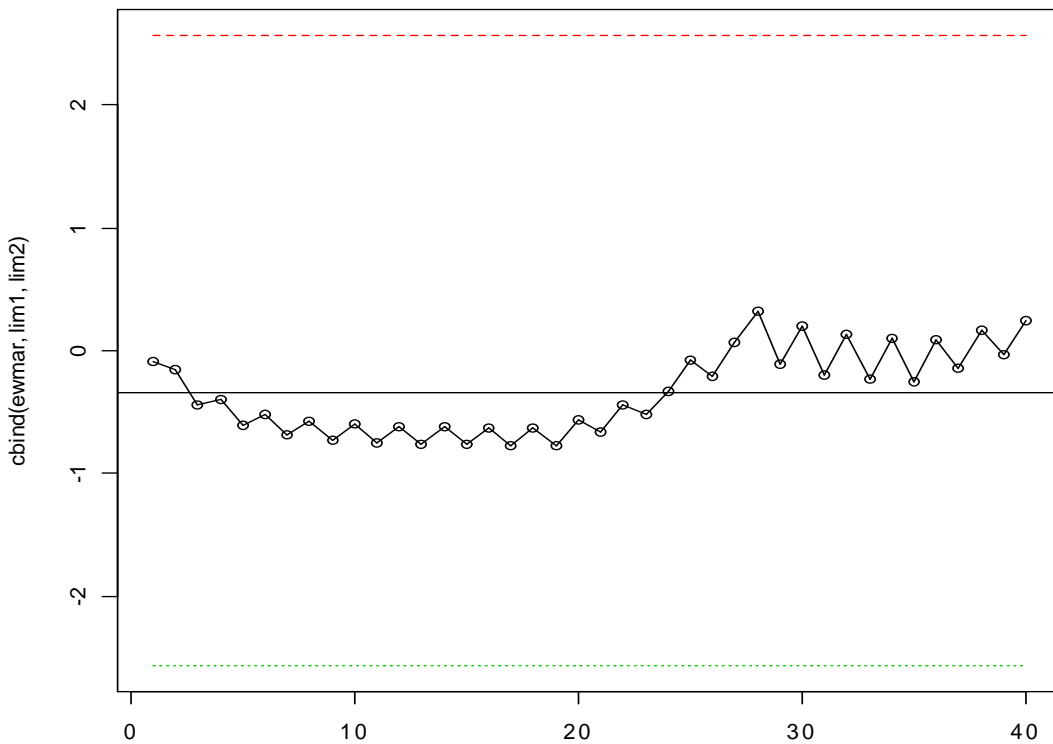
Applico la carta di Hackl e Ledolter scegliendo sempre un campione di numerosità 10 (prendo le osservazioni dalla 18 alla 27 perché più vicine alla media).



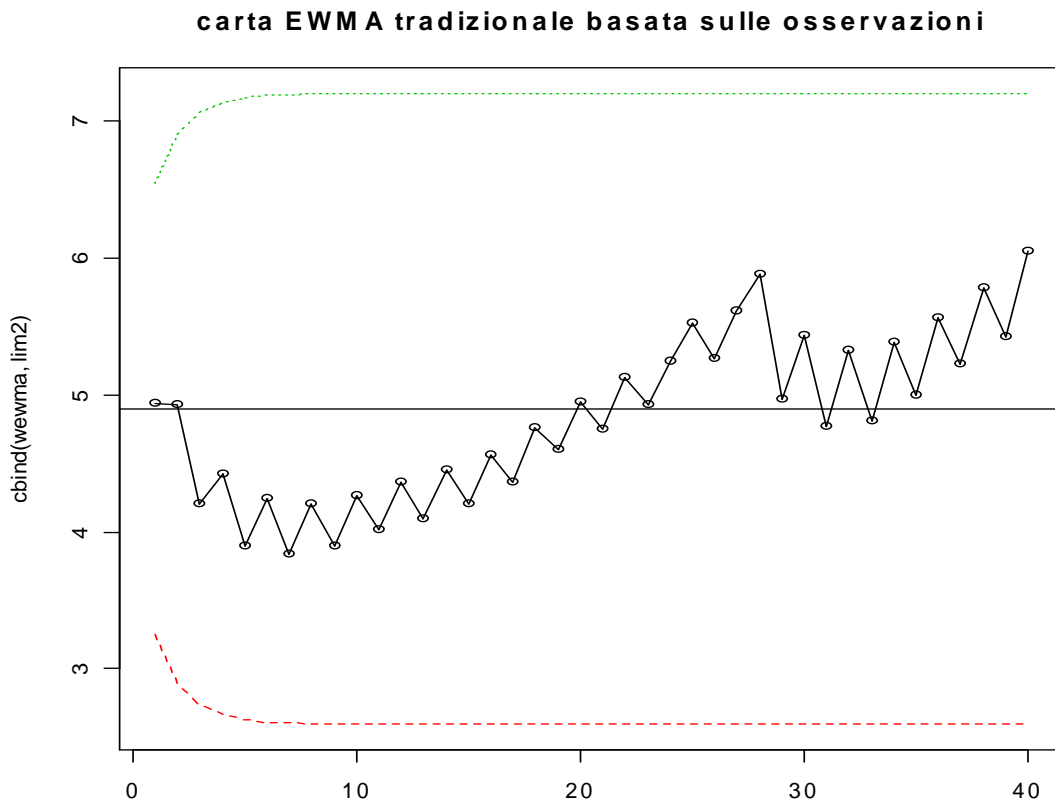
carta Shewhart per la scelta del campione



carta EWMA basata sui ranghi



Applico poi la carta EWMA basata sulle osservazioni



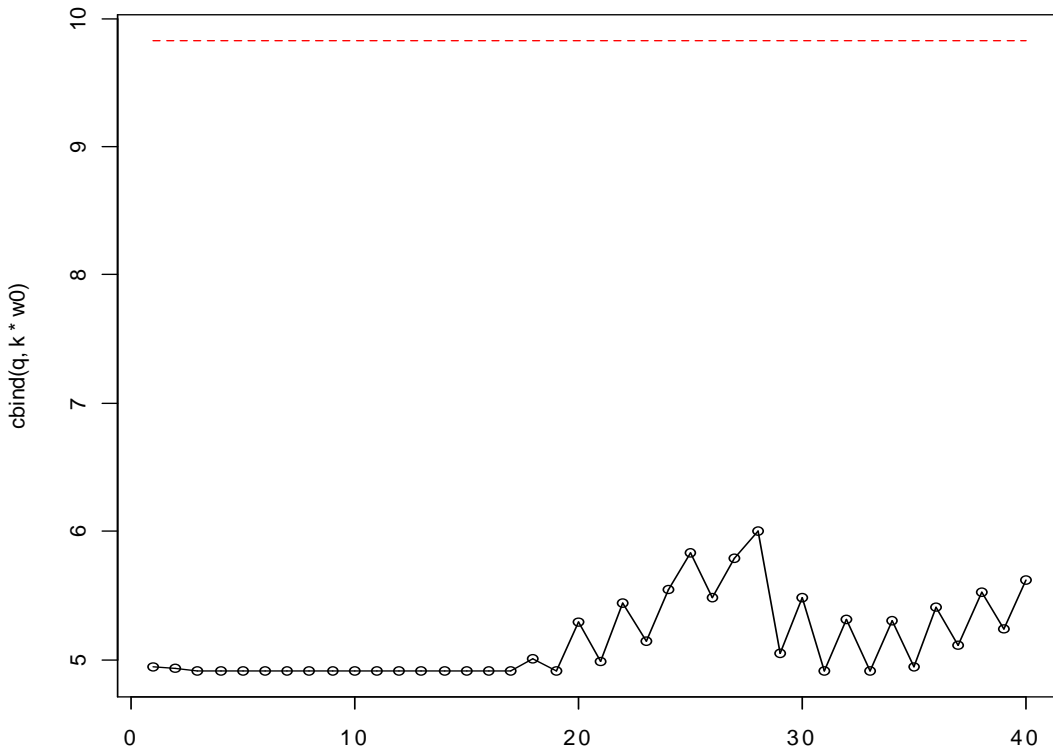
La carta EWMA tradizionale evidenzia un trend crescente che probabilmente con il passare degli istanti finirebbe per uscire dai limiti di controllo.

Per ottenere la carta EWMA per le distribuzioni di Weibull ho scelto come livello fisso di censura di 5.5.

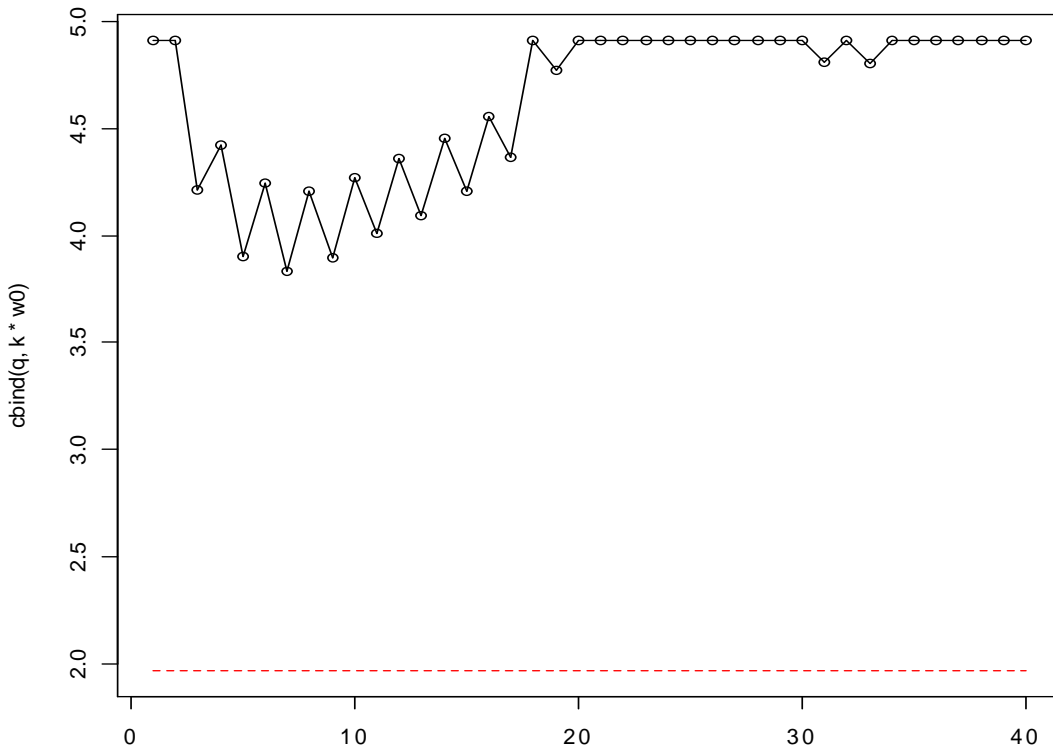
Ciò comporta una censura del 67.5% delle osservazioni.

Altro piccolo problema è che applicando ai dati non normali le trasformazioni di Zhang e Chen non si ottengono un gruppo di dati con distribuzione esponenziale di media 1. Perciò ho utilizzato la media dei dati trasformati come  $w_0$ .

**carta EWM A per distribuzione Weibull parte superiore**



**carta EWM A per distribuzione Weibull parte inferiore**

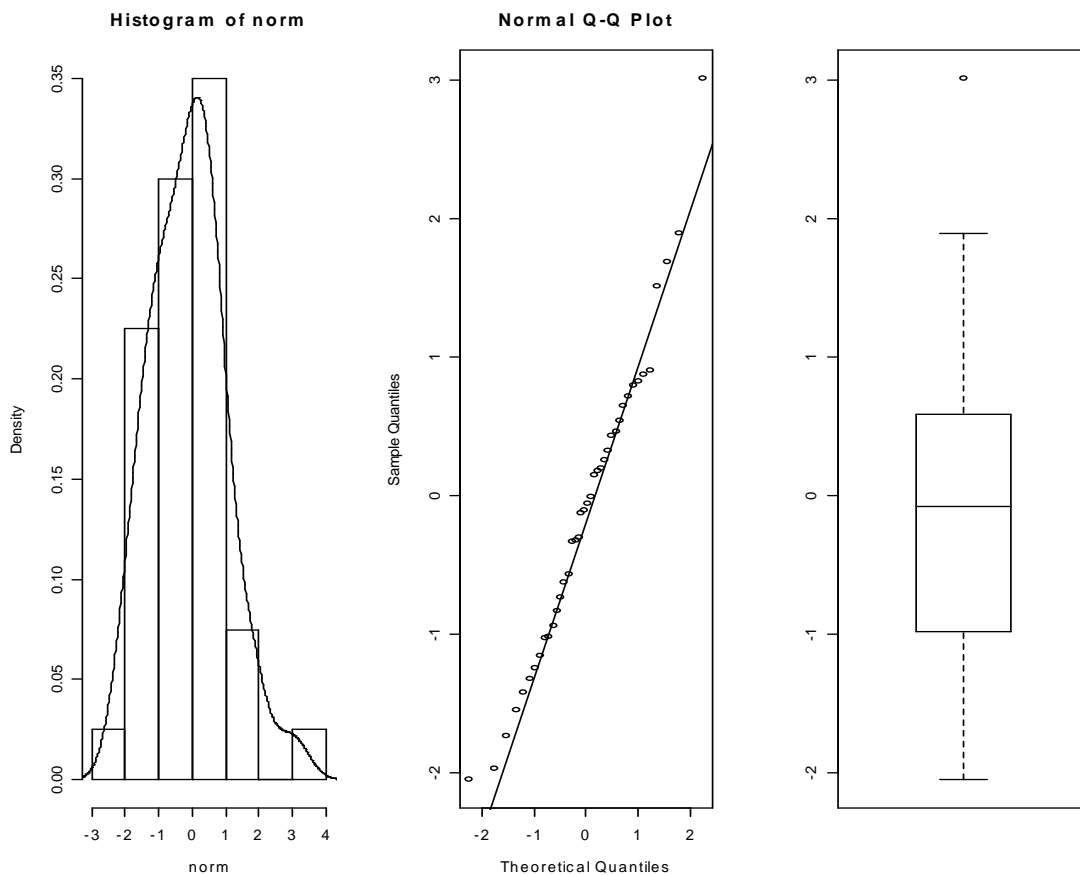


Anche queste due carte mostrano un trend crescente.

#### 4.4 Le carte EWMA con dati normali

Grazie ad R ho preso casualmente da una normale con media 0 e varianza 1 un gruppo di 40 elementi.

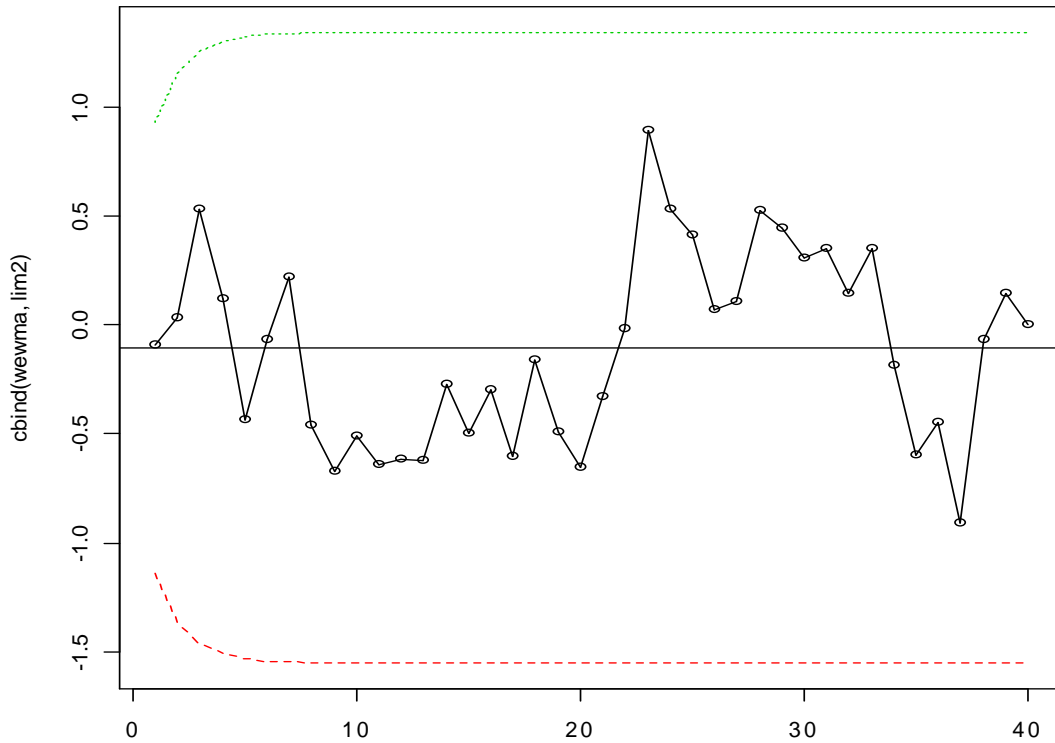
Data la bassa numerosità ne verifico la normalità con lo Shapiro test e vari grafici.



Anche se i grafici non sono particolarmente ottimi, accetto il campione con l'ulteriore conferma del test di Shapiro.

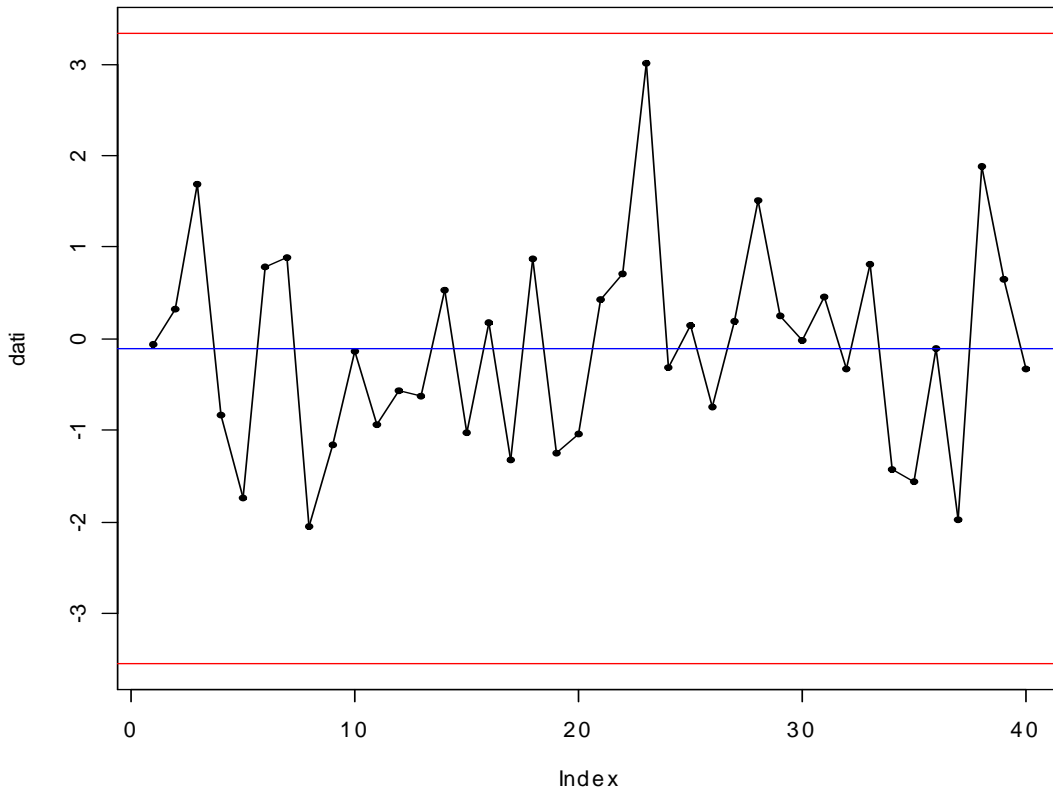
Come prima carta disegno una EWMA tradizionale, infatti mi aspetto che sia la migliore.

**carta EWMA tradizionale basata sulle osservazioni**

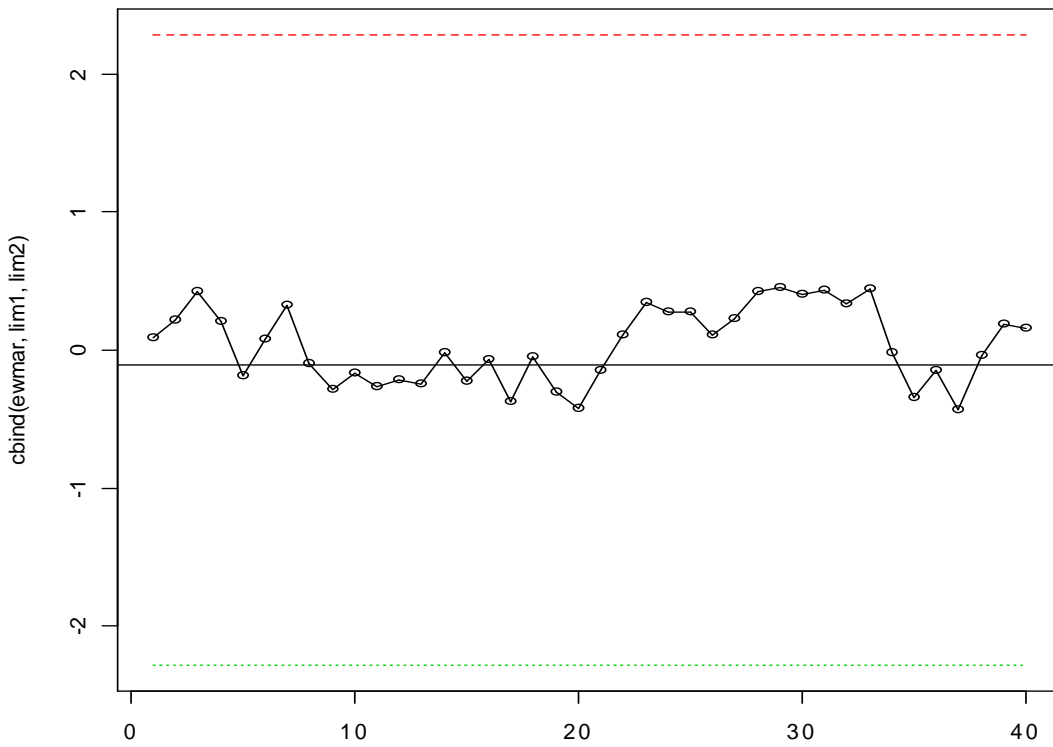


Proseguo poi con la carta EWMA basata sui ranghi, scegliendo come al solito un campione dalla carta Shewhart. Uso il campione dal 10 al 19 perché più vicino alla media.

**carta Shewhart per la scelta del campione**

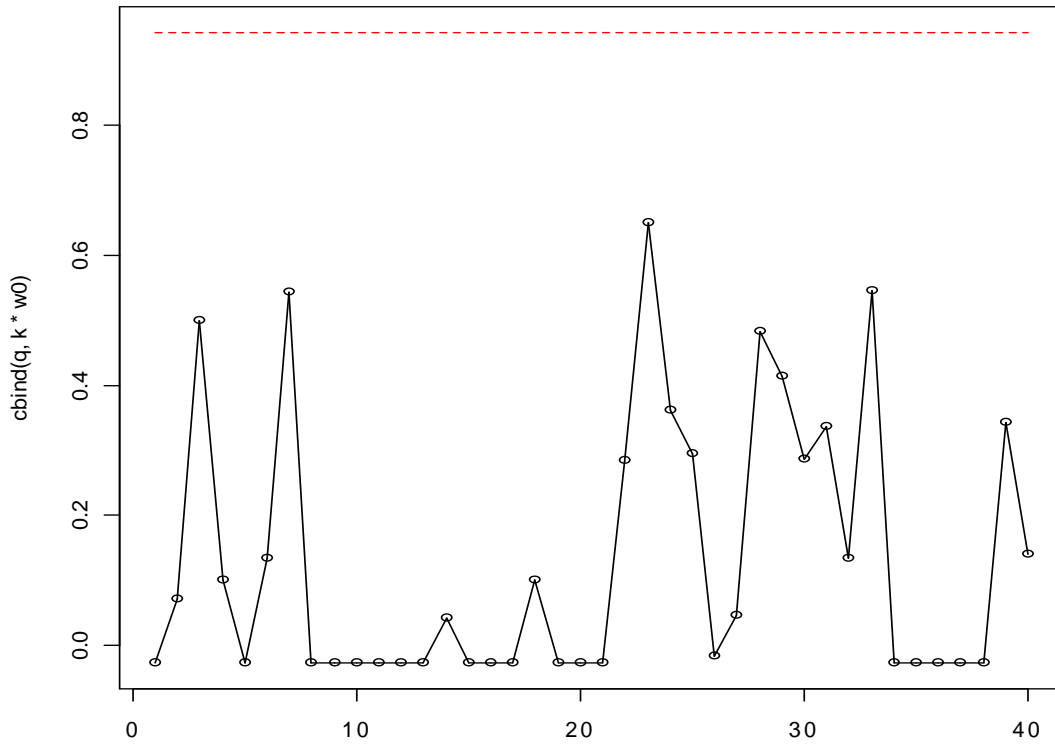


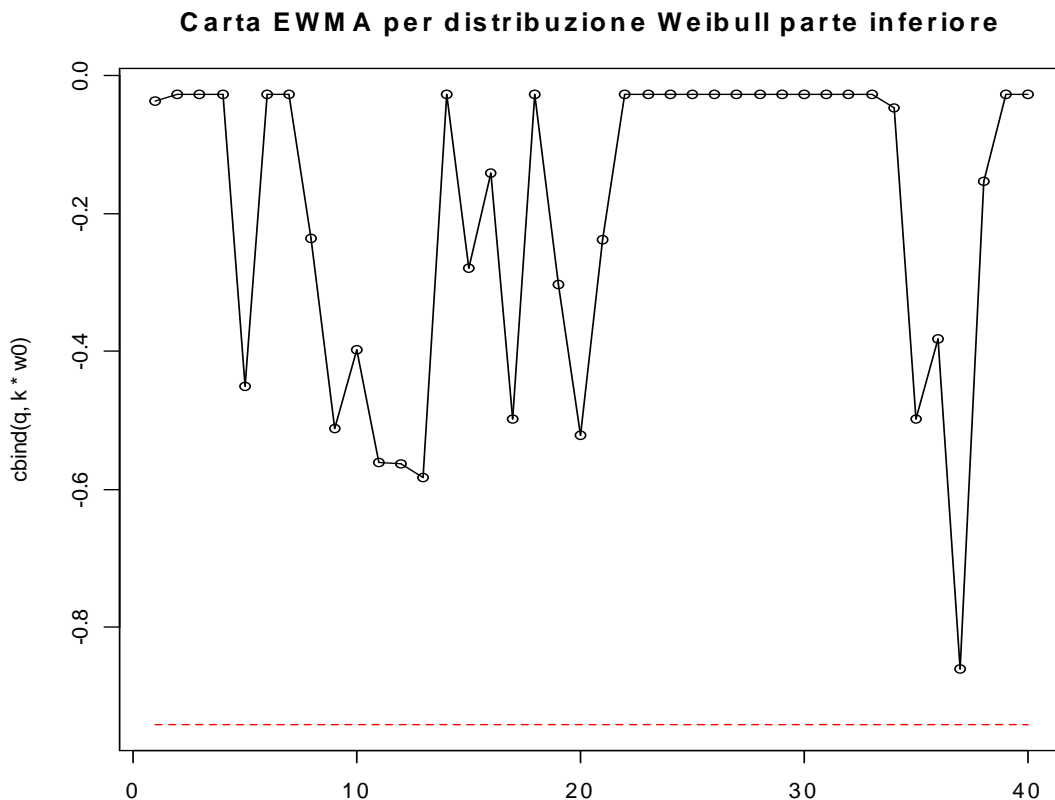
**carta EWMA basata sui ranghi**



Come precedentemente applico la carta di Zhang e Chen con un livello di censura che fisso a  $C=0.5$ . Avrò una percentuale di dati censurati pari al 72.5%.

**Carta EWM A per distribuzione Weibull parte superiore**





## 4.5 Commenti

La carta EWMA basata sui ranghi si è dimostrata molto robusta rispetto alla distribuzione di Weibull oltre che alla distribuzione normale.

La carta EWMA di Zhang e Chen, invece, perde di significato se associata ad altre distribuzioni.

Infatti se in questa carta EWMA vengono inseriti dati non adattabili ad una distribuzione di Weibull, si ha che

$$x = \left( \frac{T}{\alpha} \right)^\beta$$

non è più distribuito come una esponenziale standard di media uguale ad uno.

Allo stesso tempo cambia anche il significato dei limiti di controllo, infatti il coefficiente  $k$  non viene più moltiplicato per  $w_0=1$ .



## **BIBLIOGRAFIA**

LINGYUN ZHANG, GEMAI CHEN (2004)

“EWMA Charts for Monitoring the Mean of Censored Weibull Lifetimes”

Journal of Quality Technology

Volume 36 n°3 anno 2004

PETER HACKL, JOHANNES LEDOLTER

“A Control Chart Based on Ranks”

Journal of Quality Technology

Volume 23 n°2 anno 1991