

11 Carte UWMA e EWMA

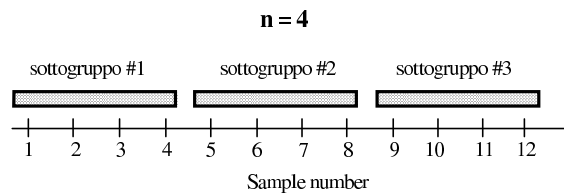
Sono carte a media mobile. UWMA: carte di controllo a media mobile uniforme; EWMA: carte di controllo a media mobile esponenziale.

Ipotesi: quando il processo è in controllo le osservazioni X_i sono variabili aleatorie indipendenti identicamente distribuite normali $\mathcal{N}(\mu, \sigma^2)$.

Per semplicità lavoriamo senza sottogruppi; l'estensione ai sottogruppi è immediata.

Abbiamo visto che le carte $\bar{X}$ fanno fatica a rilevare rapidamente piccoli spostamenti della media o della deviazione standard. Quindi per le carte $\bar{X}$ si ha un'alta probabilità β di mancato allarme. Inoltre, queste carte sono poco robuste rispetto alla gaussianità.

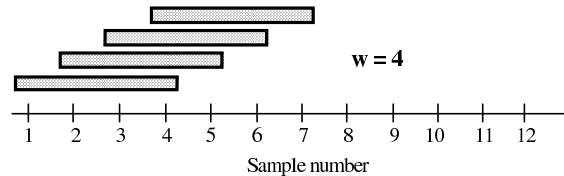
Ricordiamo che, a parità di probabilità α di falso allarme, l'uso dei sottogruppi riduce β . Un'idea potrebbe, quindi, essere quella di costruire dei sottogruppi utilizzando osservazioni successive nel tempo.



In questo modo, però, occorre troppo tempo per formare i sottogruppi. Allora un'altra idea è quella di usare una media mobile.

Carte UWMA

Sono carte di controllo a media mobile uniforme.



Si costruisce una carta di controllo per

$$M_i = \frac{X_i + X_{i-1} + \dots + X_{i-w+1}}{w}.$$

Si ha che

$$E(M_i) = \frac{E(X_i) + E(X_{i-1}) + \dots + E(X_{i-w+1})}{w} = \frac{wE(X_i)}{w} = \mu$$

poiché le X_i sono identicamente distribuite. Inoltre, visto che sono anche indipendenti:

$$\text{var}(M_i) = \frac{\text{var}(X_i) + \text{var}(X_{i-1}) + \dots + \text{var}(X_{i-w+1})}{w^2} = \frac{w\sigma^2}{w^2} = \frac{\sigma^2}{w}$$

Costruzione di una carta UWMA

- Calcolare una stima $\bar{X}$ della media μ e una stima $\hat{\sigma}_{M_i}$ della deviazione standard della media mobile M_i :

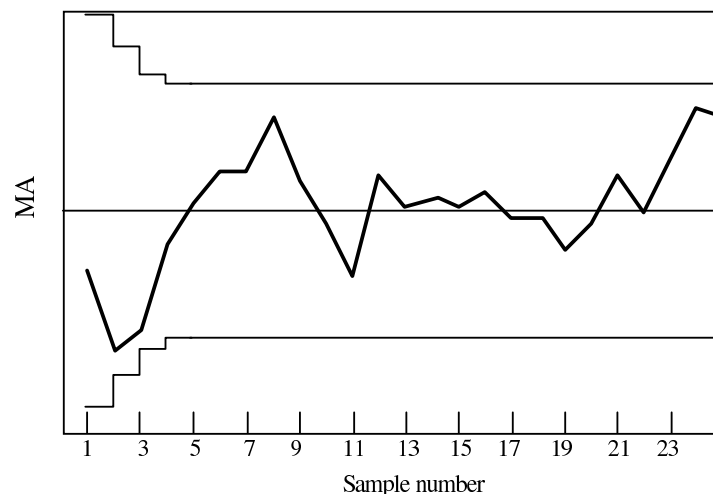
$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad ; \quad \hat{\sigma}_{M_i} = \frac{s}{\sqrt{w}}$$

- tracciare la centerline $\bar{X}$;
- calcolare e tracciare i limiti di controllo $\bar{X} \pm 3\hat{\sigma}_{M_i}$.

Per $i < w$ mancano i dati passati per poter costruire una media di w elementi. Quindi si considera

$$\begin{array}{ll} M_1 = X_1 & var(M_1) = \sigma^2 \\ M_2 = \frac{X_1 + X_2}{2} & var(M_2) = \frac{\sigma^2}{2} \\ \dots & \\ M_{w-1} = \frac{X_1 + X_2 + \dots + X_{w-1}}{w-1} & var(M_{w-1}) = \frac{\sigma^2}{w-1} \\ M_w = \frac{X_1 + X_2 + \dots + X_w}{w} & var(M_w) = \frac{\sigma^2}{w} \end{array}$$

$var(M_i)$ decresce fino a $var(M_w)$. Quindi i limiti di controllo partono larghi e poi si restringono. Più precisamente i limiti di controllo sono $\mu \pm 3\sigma/\sqrt{i}$ per $i < w$ e $\mu \pm 3\sigma/\sqrt{w}$ per $i \geq w$.

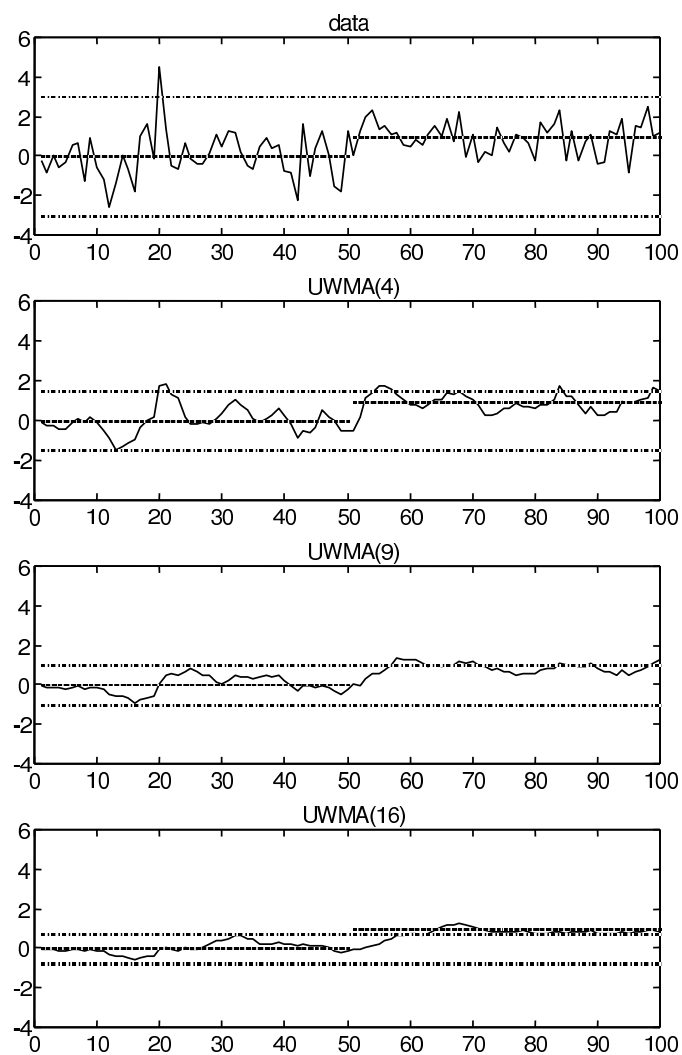


In alternativa, per $i < w$ si può utilizzare una carta $\bar{X}$.

Effetti della media mobile:

- attenua le oscillazioni puramente casuali (riduzione della varianza);

- conserva gli shift e i trend;
- se cambia improvvisamente la media, la carta di controllo UWMA impiega del tempo per andare a regime;
- introduce correlazione positiva: si possono vedere dei “run lunghi” sopra o sotto la media anche se il processo è in controllo (quindi le Western Electric Rules non sono applicabili);
- tutte le proprietà si accentuano al crescere di w .



Violazione delle ipotesi

- Gaussianità: la carta UWMA è ragionevolmente robusta purché la lunghezza della finestra sia $w \geq 4$ o 5 (per il teorema centrale del limite).
- Indipendenza: anche se le osservazioni sono correlate, al crescere di n la stima $\hat{\sigma} = s$ converge a σ (stimatore consistente). Tuttavia, il calcolo di $var(M_i)$ deve essere modificato in conseguenza alla mancanza di indipendenza.

Ad es., se $w = 2$, si ha che

$$\begin{aligned} var(M_i) &= var\left(\frac{X_i + X_{i-1}}{2}\right) = \frac{1}{4} (var(X_i) + var(X_{i-1}) + 2cov(X_i, X_{i-1})) \\ &= \frac{1}{4} (2\sigma^2 + 2cov(X_i, X_{i-1})) \neq \frac{\sigma^2}{2} \end{aligned}$$

In generale, se $cov(X_i, X_{i-k}) > 0 \quad \forall k$, le formule sottostimano $var(M_i)$ (perché uso $\sigma^2/2$ invece di $1/2(\sigma^2 + cov(X_i, X_{i-1}))$ nel caso considerato nell'esempio). Di conseguenza ho limiti di controllo troppo stretti e quindi troppi falsi allarmi.

In questi casi si può pensare, allora, di utilizzare una carta X , oppure una acceptance chart, oppure una carta per processi autocorrelati o un run to run control.

Vantaggi delle carte UWMA:

- sono meglio delle carte X per individuare piccoli spostamenti;
- i punti possono essere disegnati non appena si ottiene una nuova osservazione;
- è robusta nei confronti dell'ipotesi di gaussianità (se w non è troppo piccolo).

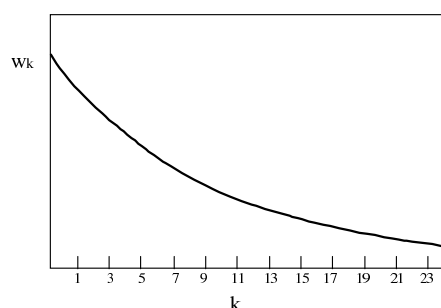
Svantaggi della carta UWMA:

- generalmente è considerata inferiore ad una carta EWMA per il rilevamento di piccoli spostamenti;
- non è robusta nei confronti dell'indipendenza e identica distribuzione, quindi è poco adatta se il processo in controllo ha cicli o trend;
- è poco adatta per segnalare discontinuità isolate (in questo caso si può usare in parallelo anche una carta X).

Carte EWMA

Nelle carte UWMA le ultime w osservazioni sono pesate allo stesso modo. Se si ha uno shift nella media del processo, all'inizio la media mobile cambia poco perché gran parte delle osservazioni precedenti sono in controllo. Quindi si ha un ritardo nel rilevare lo spostamento. L'idea è di ridurre il ritardo dando maggior peso alle ultime osservazioni. In particolare, si può usare una media mobile con finestra esponenziale:

$$Z_i = w_0 X_i + w_1 X_{i-1} + w_2 X_{i-2} + \dots$$



La carta EWMA (a media mobile esponenziale - exponentially weighted moving average) è stata introdotta da Roberts nel 1959. I punti da disegnare sulla carta sono:

$$Z_i = \lambda X_i + (1 - \lambda) Z_{i-1} \quad 0 \leq \lambda \leq 1 \quad \lambda \text{ costante}$$

e valore iniziale $Z_0 = \mu_0$. In alcuni casi, la media di dati precedenti è usata come valore iniziale della carta EWMA. In questo caso, $Z_0 = \bar{X}$.

Gli Z_i sono definiti in maniera ricorsiva:

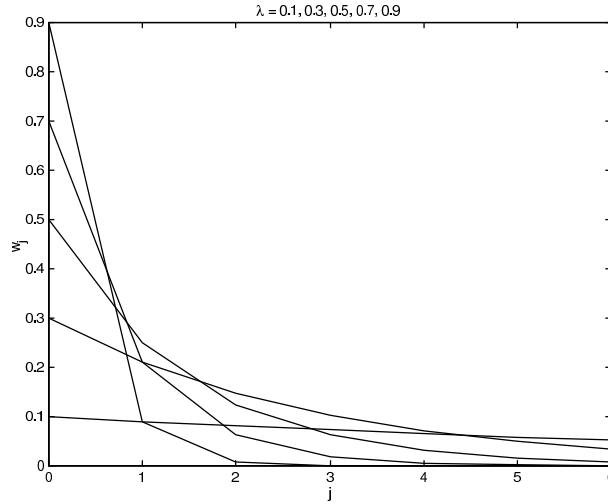
$$\begin{aligned} Z_i &= \lambda X_i + (1 - \lambda) Z_{i-1} = \lambda X_i + (1 - \lambda) [\lambda X_{i-1} + (1 - \lambda) Z_{i-2}] \\ &= \lambda X_i + \lambda(1 - \lambda) X_{i-1} + (1 - \lambda)^2 [\lambda X_{i-2} + (1 - \lambda) Z_{i-3}] \\ &= \lambda X_i + \lambda(1 - \lambda) X_{i-1} + \lambda(1 - \lambda)^2 X_{i-2} + (1 - \lambda)^3 Z_{i-3} \\ &= \dots = \lambda \sum_{j=0}^{i-1} (1 - \lambda)^j X_{i-j} + (1 - \lambda)^i Z_0 \end{aligned}$$

Si ha quindi

$$Z_i = w_0 X_i + w_1 X_{i-1} + w_2 X_{i-2} + \dots$$

con pesi $w_j = \lambda(1 - \lambda)^j$. I pesi decrescono geometricamente all'aumentare del campione.

Per questo alcune volte l'EWMA è chiamata "geometric moving average" (media mobile geometrica). Come si vede dalla figura, viene pesata di più l'osservazione corrente rispetto alle osservazioni passate.



Media e varianza delle carte EWMA

Sia $Z_0 = \mu$

$$E(Z_1) = E[\lambda X_1 + (1 - \lambda)Z_0] = \lambda E(X_1) + (1 - \lambda)\mu = \lambda\mu + (1 - \lambda)\mu = \mu$$

$$E(Z_2) = E[\lambda X_2 + (1 - \lambda)Z_1] = \lambda E(X_2) + (1 - \lambda)E(Z_1) = \lambda\mu + (1 - \lambda)\mu = \mu$$

In generale, $E(Z_i) = \mu$.

$$var(Z_1) = var[\lambda X_1 + (1 - \lambda)Z_0] = \lambda^2 var(X_1) = \lambda^2 \sigma^2$$

$$\begin{aligned} var(Z_2) &= var[\lambda X_2 + (1 - \lambda)Z_1] = \lambda^2 var(X_2) + (1 - \lambda)^2 var(Z_1) \\ &= \lambda^2 \sigma^2 + (1 - \lambda)^2 \lambda^2 \sigma^2 = \lambda^2 \sigma^2 [1 + (1 - \lambda)^2] \end{aligned}$$

$$\begin{aligned} var(Z_3) &= var[\lambda X_3 + (1 - \lambda)Z_2] = \lambda^2 var(X_3) + (1 - \lambda)^2 var(Z_2) \\ &= \lambda^2 \sigma^2 + (1 - \lambda)^2 \{ \lambda^2 \sigma^2 [1 + (1 - \lambda)^2] \} = \lambda^2 \sigma^2 [1 + (1 - \lambda)^2 + (1 - \lambda)^4] \end{aligned}$$

In generale, si ha

$$var(Z_i) = \lambda^2 \sigma^2 \sum_{j=0}^{i-1} (1 - \lambda)^{2j}$$

Ricordando che la somma geometrica è data da

$$\sum_{j=0}^n a^j = \frac{1 - a^{n+1}}{1 - a}$$

otteniamo

$$var(Z_i) = \frac{\lambda \sigma^2}{2 - \lambda} [1 - (1 - \lambda)^{2i}]$$

Osserviamo che

$$\lim_{i \rightarrow +\infty} \text{var}(Z_i) = \sigma^2 \frac{\lambda}{2 - \lambda}$$

Quindi per i abbastanza grande si può approssimare la varianza di Z_i con $\lambda\sigma^2/(2 - \lambda)$. Se $\lambda \geq 0.02$ l'approssimazione è valida già per $i \geq 5$.

Se come Z_0 si prende $\bar{X}$, allora bisognerebbe ritoccare i limiti di controllo perchè in questo caso Z_0 è una variabile aleatoria e non più una costante.

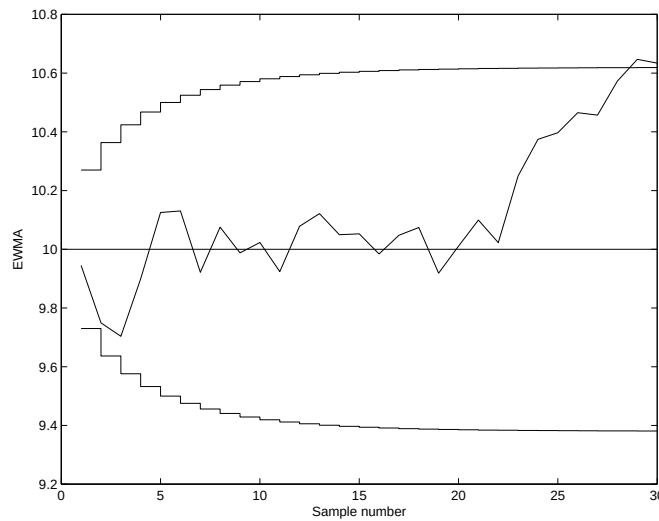
Costruzione di una carta EWMA

- Calcolare una stima $\bar{X}$ della media μ e una stima $\hat{\sigma}_{Z_i}$ della deviazione standard della media mobile Z_i :

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad ; \quad \hat{\sigma}_{Z_i} = s \sqrt{\frac{\lambda}{2 - \lambda} \left[1 - (1 - \lambda)^{2i} \right]}$$

- tracciare la centerline $\bar{X}$;
- calcolare e tracciare i limiti di controllo $\bar{X} \pm L\hat{\sigma}_{Z_i}$

$$\alpha = 0.0027 \Leftrightarrow L = 3$$



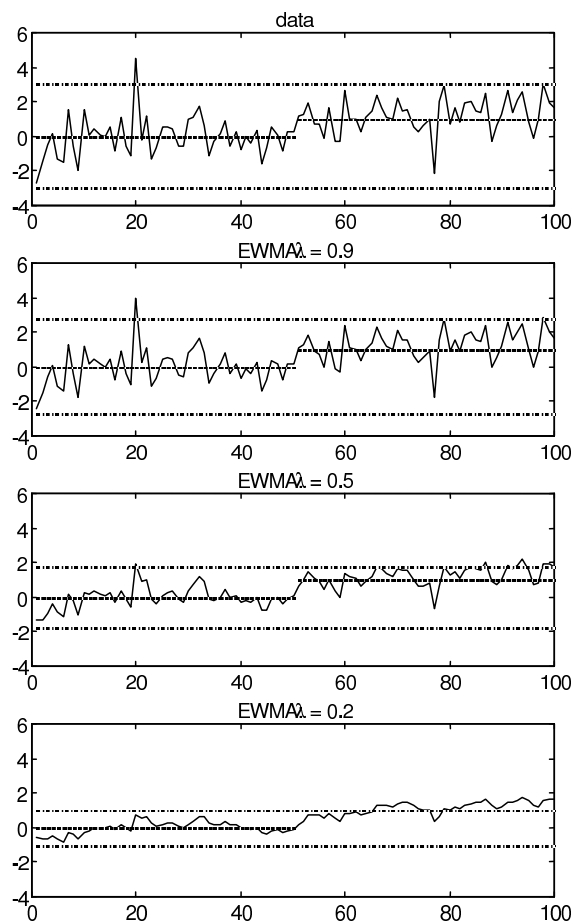
Per i grande $\hat{\sigma}_{Z_i}$ può essere approssimato con

$$\hat{\sigma}_Z = s \sqrt{\frac{\lambda}{2 - \lambda}}$$

Per i piccolo è bene usare i limiti di controllo esatti.

Effetti di EWMA:

- attenua le oscillazioni puramente casuali (riduzione della varianza);
- conserva gli shift e i trend;
- se cambia improvvisamente la media la carta EWMA impiega del tempo per andare a regime (soprattutto per λ piccolo);
- introduce correlazione positiva (si possono vedere dei run lunghi sopra o sotto la centerline anche se il processo è in controllo); allora non si possono utilizzare le Western Electric Rules;
- tutte le proprietà si accentuano al diminuire di λ .



Scelta di λ

Una regola pratica è quella di scegliere λ tanto più piccolo quanto più piccoli sono gli spostamenti che si vogliono rilevare.

Montgomery dice che valori di λ nell'intervallo $[0.05, 0.25]$ lavorano bene in pratica. Le scelte più popolari sono: $\lambda = 0.05$, $\lambda = 0.1$, $\lambda = 0.2$.

Hunter (1989) suggerisce di scegliere $\lambda = 0.4$.

Per $\lambda \rightarrow 1$, tendo alla carta $\bar{X}$.

C'è anche L da scegliere. Poiché gli Z_i sono correlati, ARL non dipende solo da α (probabilità di falso allarme), ma anche da λ . Quindi, se fisso α , ottengo diversi valori dell'ARL in funzione della scelta di λ . Allora, invece di imporre α potrei imporre l'ARL e vedere quali coppie λ , L mi permettono di ottenere quell'ARL.

Tabella 1: Coppie L , λ per avere ARL=500

L	3.054	2.998	2.962	2.814	2.615
λ	0.40	0.25	0.20	0.10	0.05

Per fare l'EWMA con sottogruppi, basta fare poche modifiche:

$$Z_i = \lambda \bar{X}_i + (1 - \lambda)Z_{i-1}$$

$$\hat{\sigma}_{Z_i} = \frac{\bar{s}}{\sqrt{n}} \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}]}$$

$$\text{Centerline: } \bar{\bar{X}} \quad ; \quad \text{control limits: } \bar{\bar{X}} \pm L\hat{\sigma}_{Z_i}$$

Montgomery sostiene che la carta EWMA fornisce una previsione di dove la media del processo sarà al prossimo tempo. Ryan sostiene invece che l'uso di una carta EWMA per la previsione e l'uso di una carta EWMA per il controllo del processo, sono a volte confusi. In realtà, EWMA fornisce una predizione della componente "shift + trend" presente nei dati solo quando si ha un fuori controllo. Infatti, se il processo è in controllo è chiaro che la previsione per la posizione futura sarà la centerline (quindi non c'è nessuna previsione), mentre se il processo è fuori controllo, la carta EWMA salva le componenti lente che sono il trend e lo shift, e quindi si riesce a prevedere dove andrà il processo se questo sta seguendo un trend.

Violazione delle ipotesi

- Gaussianità: la carta EWMA è ragionevolmente robusta purché λ sia abbastanza piccolo ($\lambda < 0.1$). Se λ è grande, la non normalità ha l'effetto di ridurre l'ARL e quindi aumenta il tasso di falsi allarmi. Questo succedeva ancora di più nelle carte individuali di Shewart in cui, anche una moderata non normalità faceva aumentare drammaticamente il tasso di falsi allarmi.

- Indipendenza: se gli X_i non sono indipendenti, allora non vale il calcolo di $var(Z_i)$ che abbiamo fatto. Se si ha correlazione positiva i limiti di controllo risultano sottostimati e quindi si hanno troppi falsi allarmi. Le cose peggiorano al diminuire di λ .

Cosa fare in questi casi? Tentare di usare λ più grande, oppure utilizzare una carta X o un'acceptance chart, oppure una carta per dati autocorrelati o un run-to-run control.

Vantaggi delle carte EWMA:

- sono meglio delle carte X per individuare piccoli spostamenti;
- i punti possono essere disegnati sulla carta non appena si ottiene una nuova osservazione;
- per λ piccolo è robusta nei confronti dell'ipotesi di gaussianità.

Svantaggi delle carte EWMA:

- non è robusta nei confronti di indipendenza e identica distribuzione; quindi è poco adatta se il processo in controllo ha cicli o trend;
- è poco adatta per segnalare discontinuità isolate; come possibile rimedio si può usare in parallelo anche una carta X (volendo le due carte possono essere sovrapposte).

In conclusione: l'EWMA funziona bene per rilevare piccoli spostamenti, ma non reagisce ai grandi spostamenti così rapidamente come una carta di Shewart. Un buon modo per migliorare la sensitività della procedura di controllo agli spostamenti grandi senza sacrificare l'abilità di rilevare spostamenti piccoli, è quello di combinare una carta di Shewart con una EWMA.

Esercizio

Per la caratteristica di un processo sono note media e deviazione standard: $\mu = 100$, $\sigma = 0.2$. Calcolare la centerline e i limiti di controllo di regime della carta UWMA con $w = 5$.

La centerline è data da $\bar{X} = 100$. I limiti di controllo di regime sono dati da: $\mu \pm 3\hat{\sigma}_{M_i}$ dove

$$\hat{\sigma}_{M_i} = \frac{s}{\sqrt{w}} = \frac{0.2}{\sqrt{5}} = 0.089$$

Segue che

$$LCL = \mu - 3\hat{\sigma}_{M_i} = 100 - 3 \cdot 0.089 = 99.33$$

$$UCL = \mu + 3\hat{\sigma}_{M_i} = 100 + 3 \cdot 0.089 = 100.267$$

Calcolare i limiti di controllo di una carta EWMA per un campione molto grande, utilizzando come λ e L i valori: $\lambda = 0.1$ e $L = 3$.

I limiti di controllo per un campione molto grande possono essere approssimati con $\bar{X} \pm L\hat{\sigma}_Z$ dove

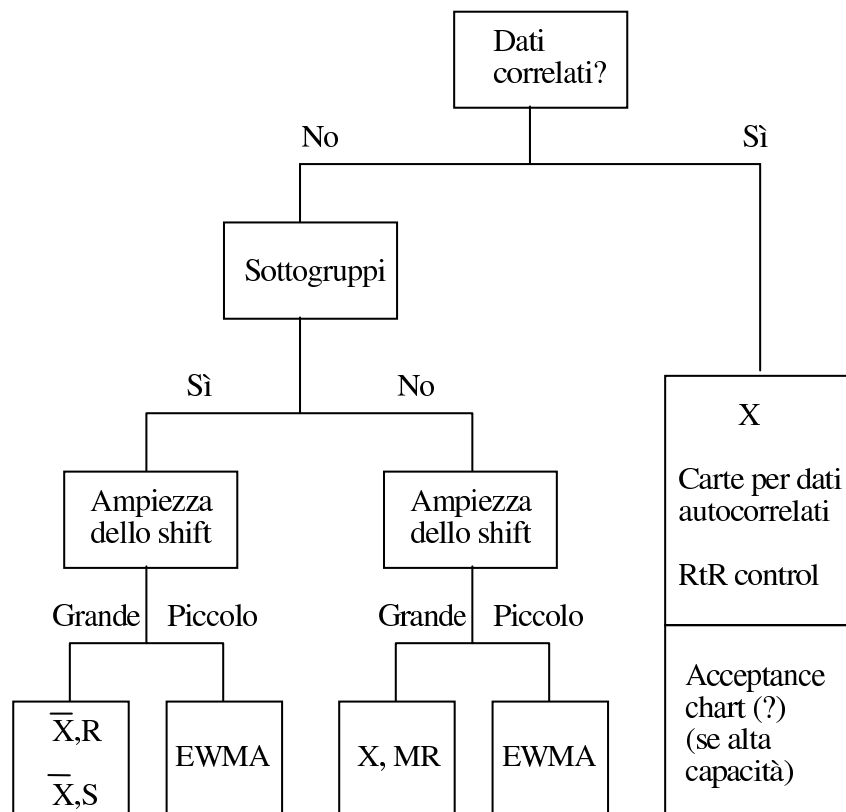
$$\hat{\sigma}_Z = s\sqrt{\frac{\lambda}{2-\lambda}} = 0.2\sqrt{\frac{0.1}{2-0.1}} = 0.2 \cdot 0.23 = 0.046$$

Segue che i limiti di controllo richiesti sono

$$LCL = \bar{X} - 3\hat{\sigma}_Z = 100 - 3 \cdot 0.046 = 99.862$$

$$UCL = \bar{X} + 3\hat{\sigma}_Z = 100 + 3 \cdot 0.046 = 100.138.$$

Linee guida per la scelta della carta di controllo (caso univariato)



12 Carte per attributi

Molte caratteristiche di qualità non possono essere rappresentate numericamente. In questi casi allora si controlla se il prodotto è conforme o meno alle specifiche, relativamente alla caratteristica di qualità in esame. In passato si parlava di unità difettose e difetti, la terminologia più recente (normativa ANSI/ASQC A1 - 1987) invece parla di unità non conformi e non conformità.

Le carte per attributi sono generalmente meno informative delle carte per variabili perché, tipicamente, c'è più informazione in una misura numerica che nella classificazione delle unità in conformi o non conformi. Comunque, le carte per attributi hanno importanti applicazioni. Sono particolarmente utili nelle industrie di servizi e nei settori manifatturieri perché molte delle caratteristiche di qualità che si trovano in questi settori non sono facilmente misurabili su scala numerica.

Ci sono diversi tipi di carte per attributi:

- quando si controlla il numero di prodotti non conformi $\Rightarrow$ carte p e np ;
- quando si controlla il numero di non conformità nei prodotti $\Rightarrow$ carte c e u .

Osservazione: è meglio prevenire prodotti non conformi utilizzando carte per variabili (ad es. $\bar{X}$ o X) che controllare il numero di prodotti non conformi con carte p e np .

Nelle carte p e np si considera la frazione di prodotti non conformi che è data dal rapporto tra il numero di prodotti non conformi e il numero totale di prodotti della popolazione.

Es. Si vuole controllare il macchinario che produce le scatole di cartone che devono contenere succo di frutta. Il macchinario di confezionamento è ritenuto in controllo se la scatola prodotta è sigillata bene e non perde liquido (potrebbe perdere dai lati delle “cuciture” oppure intorno al tappo). In questo caso per verificare che la macchina è in controllo si esamina la frazione dei prodotti non conformi.

Nelle carte c e u invece si considerano il numero di non conformità in una unità o il numero medio di non conformità per unità.

Es. Si vogliono controllare dei tessuti e si esamina l'occorrenza di difetti per 50 m^2 di tessuto. Una unità è data da 50 m^2 di tessuto e si osservano i difetti che ci sono in questa unità.

L'ipotesi su cui si basano queste carte è che il numero di non conformità in campioni di dimensione costante è una variabile aleatoria con distribuzione di Poisson.

Processi di Poisson spaziali

Consideriamo dei punti distribuiti casualmente su una superficie, come, ad es., le gocce di pioggia su una mattonella in un minuto, il numero di stelle osservate in una piccola area di cielo, le particelle su un wafer. Il classico modello che descrive la distribuzione di questi punti è il modello di Poisson. Quindi queste sono variabili aleatorie con distribuzione di Poisson.

Pensiamo di dividere una zona di piano in regioni ΔS_k , di uguale area ΔS , e indichiamo con $N(\Delta S)$ il numero di punti che cadono nell'area ΔS .

	...	...	...	
...	ΔS_1	ΔS_2	ΔS_3	...
...	ΔS_4	ΔS_5	ΔS_6	...
...	ΔS_7	ΔS_8	ΔS_9	...
	...	...	...	

Il processo $\{N(\Delta S_k), k = 0, 1, 2, \dots\}$ è detto processo di Poisson spaziale omogeneo se:

1. comunque presa una regione ΔS , la variabile aleatoria $N(\Delta S)$ ha distribuzione di Poisson con parametro $\lambda|\Delta S|$ (dove $|\Delta S|$ è la dimensione di ΔS , che può essere, una lunghezza, un'area, un volume e $\lambda > 0$ è una costante):

$$P(N(\Delta S) = j) = e^{-\lambda|\Delta S|} \frac{(\lambda|\Delta S|)^j}{j!} \quad j = 0, 1, 2, \dots$$

2. comunque presa una sequenza finita di regioni $\{\Delta S_1, \Delta S_2, \dots, \Delta S_n\}$ disgiunte, allora le variabili aleatorie $N(\Delta S_1), N(\Delta S_2), \dots, N(\Delta S_n)$ sono indipendenti.

Una definizione alternativa è data da:

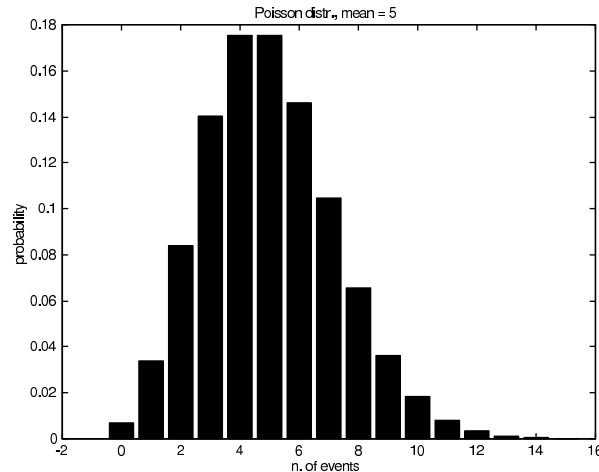
1. la distribuzione di probabilità di $N(\Delta S)$ dipende dalla dimensione di ΔS ;
2. se $\Delta S_1, \Delta S_2, \dots, \Delta S_n$ sono regioni disgiunte, allora le variabili aleatorie $N(\Delta S_1), N(\Delta S_2), \dots, N(\Delta S_n)$ sono indipendenti e $N(\Delta S_1 \cup \Delta S_2 \cup \dots \cup \Delta S_n) = N(\Delta S_1) + N(\Delta S_2) + \dots + N(\Delta S_n)$;
3. $P(N(\Delta S) = 1) = \lambda|\Delta S| + o(|\Delta S|)$ per $|\Delta S| \rightarrow 0$;
4. $P(N(\Delta S) > 1) = o(|\Delta S|)$ per $|\Delta S| \rightarrow 0$.

La condizione 1 afferma che la distribuzione di probabilità di $N(\Delta S)$ non dipende dalla forma o dalla posizione della regione ΔS , ma solo dalla sua dimensione. La condizione 2 afferma che, sapere che in una certa regione c'è un evento, nulla dice sulla probabilità che nelle vicinanze ci sia un altro evento. La condizione 4 preclude che, quando le regioni ΔS_k sono piccole, ci siano due punti nella stessa area.

Come interpretazione possiamo dire che, seguendo un processo di Poisson, i punti sono sparpagliati in modo casuale nel piano senza preferenza per alcune regioni e senza correlazioni.

Nel piano, la probabilità che in una regione di area S vi siano k eventi è data da:

$$P(N(S) = k) = e^{-\lambda S} \frac{(\lambda S)^k}{k!}$$



Segue che:

- $E(N) = \lambda S$;
- $var(N) = \lambda S$;
- la distribuzione di Poisson può essere approssimata con una gaussiana quando $\lambda S \rightarrow \infty$ (l'approssimazione è ragionevole per $E(N) > 5$).

Es. Bombardamento di Londra durante la II Guerra Mondiale: sono sganciate 3.729 bombe/ Km^2 . Assumendo una distribuzione di Poisson, calcolare la probabilità che 0, 1 bombe colpiscano una zona di 1 km^2 e la probabilità che più di una bomba colpisca una zona di 1 km^2 .

In questo caso $\lambda = 3.729$ e $S = 1$, quindi $\lambda S = 3.729$.

$$P(N(1) = 0) = e^{-3.729} \frac{(3.729)^0}{0!} = 0.024$$

$$P(N(1) = 1) = e^{-3.729} \frac{(3.729)^1}{1!} = 0.0895$$

$$P(N(1) > 1) = 1 - P(N(1) = 0) - P(N(1) = 1) = 0.8865.$$

Carta c

L'obiettivo di una carta c è quello di tenere sotto controllo il numero di non conformità. I punti da disegnare sulla carta sono

C_i = n. di non conformità nell' i -ma ispezione.

L'ipotesi di fondo è che le variabili aleatorie C_i sono indipendenti e identicamente distribuite di Poisson con $E(C_i) = \mu$.

Osserviamo che in questo caso non ha senso parlare di sottogruppi. Se ispeziono 5 oggetti invece di 1, il numero di non conformità è ancora una variabile aleatoria di Poisson ma con media 5 volte più grande (perché la superficie è quintuplicata). In questo caso l'“unità ispezionata” è costituita da 5 oggetti invece che da 1. La procedura per la costruzione della carta non cambia se l'“unità ispezionata” è costituita da 5 oggetti oppure da 1 ma, a parità di α , diminuisce la probabilità β di mancato allarme.

Costruzione di una carta c

- Calcolare una stima $\bar{c}$ della media μ

$$\bar{c} = \frac{\sum_{i=1}^n C_i}{n}$$

dove n è il numero di ispezioni;

- tracciare la centerline $\bar{c}$;
- calcolare e tracciare i limiti di controllo; i limiti di controllo si ricavano utilizzando la distribuzione di Poisson. Quando μ è grande si può approssimare la distribuzione di Poisson con la gaussiana. In questo caso i limiti di controllo sono $\bar{c} \pm 3\sqrt{\bar{c}}$. Per μ piccolo la Poisson è fortemente asimmetrica e quindi è bene calcolare i limiti di controllo esatti e non utilizzare l'approssimazione normale.

Come verificare la Poissonianità

Sia X_i il numero di eventi o particelle nell' i -ma ispezione. Se X è Poissoniana, dato che $E(X) = \text{var}(X)$ mi aspetto che

$$V := \frac{S^2}{\bar{X}} \simeq 1.$$

Se V è molto diverso da 1, allora respingo l'ipotesi che gli eventi sono Poissoniani.

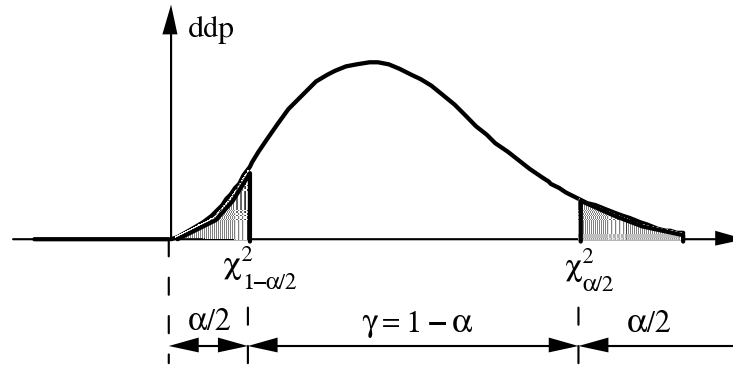
Cosa si intende per molto diverso?

Se $E(X)$ è sufficientemente grande (cioè > 5), allora

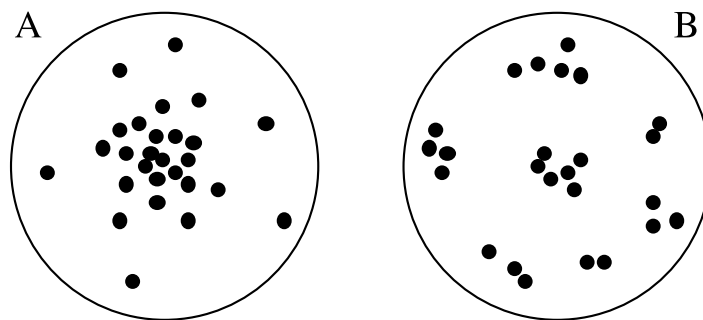
$$\xi = \frac{(n-1)S^2}{\bar{X}} \sim \chi_{n-1}^2$$

dove n è il numero di ispezioni.

A questo punto si applica il test di dispersione: fissato α si trovano $\chi_{\alpha/2, n-1}^2$ e $\chi_{1-\alpha/2, n-1}^2$. Se $\chi_{1-\alpha/2, n-1}^2 < \xi < \chi_{\alpha/2, n-1}^2$, allora si accetta l'ipotesi che gli X_i siano distribuiti di Poisson, altrimenti si rifiuta l'ipotesi di Poissonianità.



Violazione della Poissonianità



Nel caso A viene violata l'uniformità: ci sono regioni in cui la densità di eventi è maggiore. In questo caso l'idea è quella di prendere λ dipendente dalla posizione: $\lambda = \lambda(x, y)$. Otteniamo così un processo di Poisson non omogeneo.

Nel caso B viene violata l'indipendenza: se c'è un evento in un punto è più probabile che ce ne siano altri nelle vicinanze (clusters). Nel nostro caso, vuol dire che se c'è un difetto in un'area, è probabile che ci siano altri difetti nelle vicinanze. Nel caso B si può usare un modello con 2 distribuzioni: una distribuzione per il numero di cluster e una per il numero di eventi per cluster.

Se è violata la Poissonianità, allora si avranno limiti di controllo troppo stretti o troppo larghi e, di conseguenza, troppi falsi allarmi o mancati allarmi.

Es. Si considera il numero di non conformità riscontrate su 26 campioni costituiti da 100 circuiti stampati (quindi l'unità ispezionata è costituita da 100 circuiti).

numero campione	numero di non conformità	numero campione	numero di non conformità
1	21	14	19
2	24	15	10
3	16	16	17
4	12	17	13
5	15	18	22
6	5	19	18
7	28	20	39
8	20	21	30
9	31	22	24
10	25	23	16
11	20	24	19
12	24	25	17
13	16	26	15

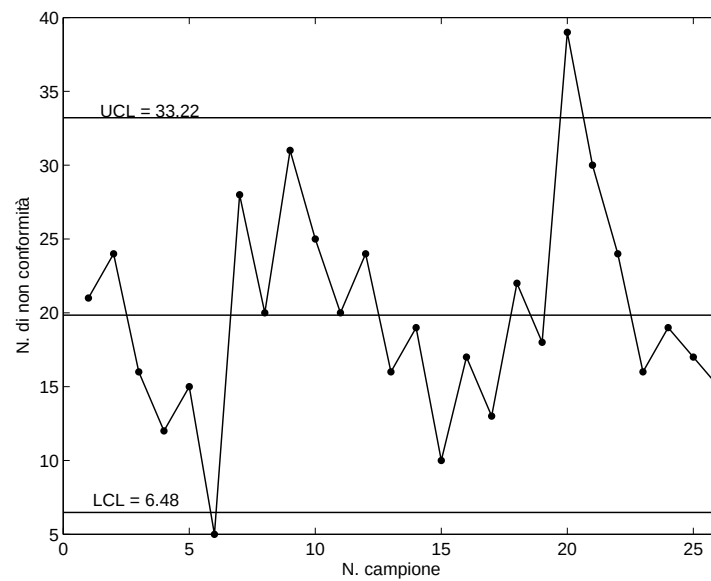
Il totale delle non conformità per i 26 campioni è 516. Segue che

$$\bar{c} = \frac{\sum_{i=1}^n c_i}{n} = \frac{516}{26} = 19.85$$

Poiché la stima della media è grande, allora si può utilizzare l'approssimazione della Poisson con la gaussiana. Quindi i limiti di controllo sono:

$$UCL = \bar{c} + 3\sqrt{\bar{c}} = 19.85 + 3\sqrt{19.85} = 33.22$$

$$LCL = \bar{c} - 3\sqrt{\bar{c}} = 19.85 - 3\sqrt{19.85} = 6.48$$



Carta u

Nell'esempio visto si calcola il numero di non conformità per ogni "unità ispezionata". Questa unità è scelta per semplicità nella raccolta dati. Non c'è però nessuna ragione per restringersi ad una singola unità. Infatti, spesso si preferisce usare più unità nel campione aumentando così l'area in esame per il conteggio delle non conformità. In questo caso, si può utilizzare una carta di controllo basata sul numero medio di non conformità per unità. La carta di controllo per non conformità per unità si dice carta u .

In generale, se C_i è il numero di non conformità all' i -ma ispezione, nella carta u si disegneranno i punti

$$U_i = \frac{C_i}{r}$$

dove r è il numero di unità ispezionate in ogni ispezione oppure è l'area ispezionata.

- se r è l'area ispezionata $\Rightarrow u$ è una stima della densità λ delle non conformità;
- se r è il numero di unità ispezionate $\Rightarrow u$ è una stima del numero medio di non conformità per unità.

Anche qui si suppone che le C_i siano variabili aleatorie indipendenti e identicamente distribuite di Poisson con $E(C_i) = \mu$.

Costruzione di una carta u

- Calcolare una stima $\bar{u}$ della media μ/r :

$$\bar{u} = \frac{\bar{c}}{r} = \frac{\sum_{i=1}^n c_i}{n \cdot r} = \frac{\sum_{i=1}^n u_i}{n}$$

- tracciare la centerline $\bar{u}$;
- calcolare e tracciare i limiti di controllo. I limiti di controllo possono essere ricavati da quelli di una carta c :

$$LCL_u = \frac{LCL_c}{r} \quad ; \quad UCL_u = \frac{UCL_c}{r}$$

Nel caso in cui valga l'approssimazione della Poisson con la gaussiana, tali limiti diventano:

$$LCL_u = \bar{u} - 3\sqrt{\frac{\bar{u}}{r}} \quad ; \quad UCL_u = \bar{u} + 3\sqrt{\frac{\bar{u}}{r}}.$$

La carta u è utile quando il numero di oggetti ispezionati (o l'area ispezionata) varia nel tempo.

Ad esempio, nell'ispezione di rotoli di tessuto o rotoli di carta, spesso la lunghezza dei rotoli cambia. Se utilizziamo una carta c per le non conformità, allora sia la centerline che i limiti di controllo cambiano con la dimensione del campione, e la carta è di difficile

interpretazione. In questi casi è corretto usare una carta u . La carta u ha la centerline costante e i limiti di controllo varieranno con l'ampiezza del campione.

Es. Un produttore di PC vuole utilizzare una carta di controllo per non conformità per unità sulla linea di assemblaggio finale. La dimensione del campione è posta uguale a 5 PC.

numero campione	dimensione campione	numero totale di non conformità	numero medio di non conformità per unità
1	5	10	2.0
2	5	12	2.4
3	5	8	1.6
4	5	14	2.8
5	5	10	2.0
6	5	16	3.2
7	5	11	2.2
8	5	7	1.4
9	5	10	2.0
10	5	15	3.0
11	5	9	1.8
12	5	5	1.0
13	5	7	1.4
14	5	11	2.2
15	5	12	2.4
16	5	6	1.2
17	5	8	1.6
18	5	10	2.0
19	5	7	1.4
20	5	5	1.0
		193	38.6

La centerline è data da:

$$\bar{u} = \frac{\bar{c}}{r} = \frac{\sum_{i=1}^n c_i/n}{r} = \frac{193/20}{5} = 1.93$$

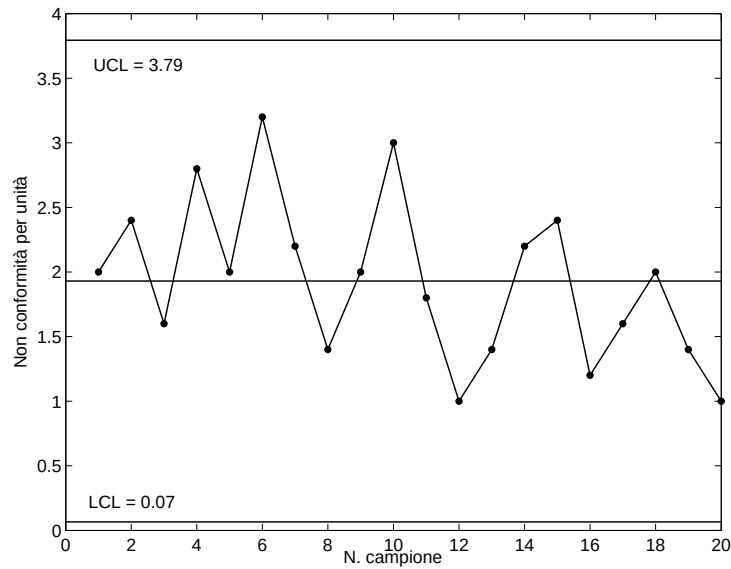
oppure da

$$\bar{u} = \frac{\sum_{i=1}^n u_i}{n} = \frac{38.6}{20} = 1.93$$

Poiché la stima della media è grande ($193/20=9.65$), possiamo approssimare la Poisson con la normale. Pertanto, i limiti di controllo sono dati da:

$$UCL = \bar{u} + 3\sqrt{\frac{\bar{u}}{r}} = 1.93 + 3\sqrt{\frac{1.93}{5}} = 3.79$$

$$LCL = \bar{u} - 3\sqrt{\frac{\bar{u}}{r}} = 1.93 - 3\sqrt{\frac{1.93}{5}} = 0.07.$$



Es. In una fabbrica tessile vengono controllati dei rotoli di tessuto, analizzando il numero di non conformità per $50 m^2$ di tessuto.

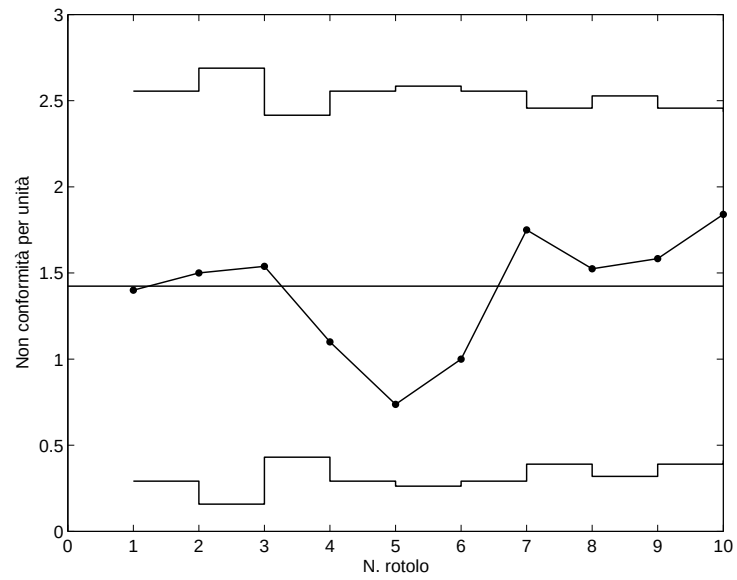
rotolo numero	numero di m^2	numero totale di non conformità	numero di unità ispezionate nel rotolo	numero medio di non conformità per unità ispezionata
1	500	14	10.0	1.40
2	400	12	8.0	1.50
3	650	20	13.0	1.54
4	500	11	10.0	1.10
5	475	7	9.5	0.74
6	500	10	10.0	1.00
7	600	21	12.0	1.75
8	525	16	10.5	1.52
9	600	19	12.0	1.58
10	625	23	12.5	1.84
		153	107.5	13.97

Il numero di m^2 per ogni rotolo non è lo stesso, quindi la dimensione di ogni campione cambia (essendo l'unità costituita da $50 m^2$). La centerline è data dal numero medio di non conformità per unità ispezionata. Quindi

$$\bar{u} = \frac{153}{107.5} = 1.42$$

($\bar{u}$ è dato dal rapporto tra il numero di non conformità osservate diviso il numero di unità ispezionate). I limiti di controllo variano al variare della dimensione del campione.

rotolo			
numero	r_i	$UCL = \bar{u} + 3\sqrt{\bar{u}/r_i}$	$LCL = \bar{u} - 3\sqrt{\bar{u}/r_i}$
1	10.0	2.55	0.29
2	8.0	2.68	0.16
3	13.0	2.41	0.43
4	10.0	2.55	0.29
5	9.5	2.58	0.26
6	10.0	2.55	0.29
7	12.0	2.45	0.39
8	10.5	2.52	0.32
9	12.0	2.45	0.39
10	12.5	2.43	0.41



Esercizio. Vengono ispezionati 26 cavi telefonici di 1000 m ciascuno. La stima del numero medio di non conformità per ogni cavo è 12.3, mentre la stima della varianza del numero di non conformità è 29.16. Valutare, mediante il test di dispersione (con significatività $\alpha = 0.05$) se l'ipotesi di distribuzione Poissoniana deve essere respinta.

Sappiamo che

$$\frac{(n-1)S^2}{\bar{X}} \sim \chi_{n-1}^2$$

In questo caso è $n = 26$, $S^2 = 29.16$, $\bar{X} = 12.3$ e $\alpha = 0.05$. Dobbiamo trovare i due limiti $\chi_{\alpha/2, n-1}^2$ e $\chi_{1-\alpha/2, n-1}^2$. Per far questo usiamo le tavole del χ^2 .

Valori di χ_{α}^2

n	$\alpha=0.975$	$\alpha=0.95$	$\alpha=0.05$	$\alpha=0.025$	n	$\alpha=0.975$	$\alpha=0.95$	$\alpha=0.05$	$\alpha=0.025$
1	0.00	0.00	3.84	5.02	16	6.91	7.96	26.30	28.85
2	0.05	0.10	5.99	7.38	17	7.56	8.67	27.59	30.19
3	0.22	0.35	7.81	9.35	18	8.23	9.39	28.87	31.53
4	0.48	0.71	9.49	11.14	19	8.91	10.12	30.14	32.85
5	0.83	1.15	11.07	12.38	20	9.59	10.85	31.41	34.17
6	1.24	1.64	12.59	14.45	25	13.12	14.61	37.65	40.65
7	1.69	2.17	14.07	16.01	30	16.79	18.49	43.77	46.98
8	2.18	2.73	15.51	17.53	40	24.43	26.51	55.76	59.34
9	2.70	3.33	16.92	19.02	50	32.36	34.76	67.50	71.42
10	3.25	3.94	18.31	20.48	60	40.48	43.19	79.08	83.30
11	3.82	4.57	19.68	21.92	70	48.76	51.74	90.53	95.02
12	4.40	5.23	21.03	23.34	80	57.15	60.39	101.88	106.63
13	5.01	5.89	22.36	24.74	90	65.65	69.13	113.14	118.14
14	5.63	6.57	23.68	26.12	100	74.22	77.93	124.34	129.50
15	6.27	7.26	25.00	27.49					

Dalla tabella troviamo che $\chi_{0.025, 25}^2 = 40.65$ e $\chi_{0.975, 25}^2 = 13.12$.

$$\frac{(n-1)S^2}{\bar{X}} = \frac{25 \cdot 29.16}{12.3} = 59.27 \notin [13.12, 40.65]$$

quindi rifiuto l'ipotesi di Poissonianità.

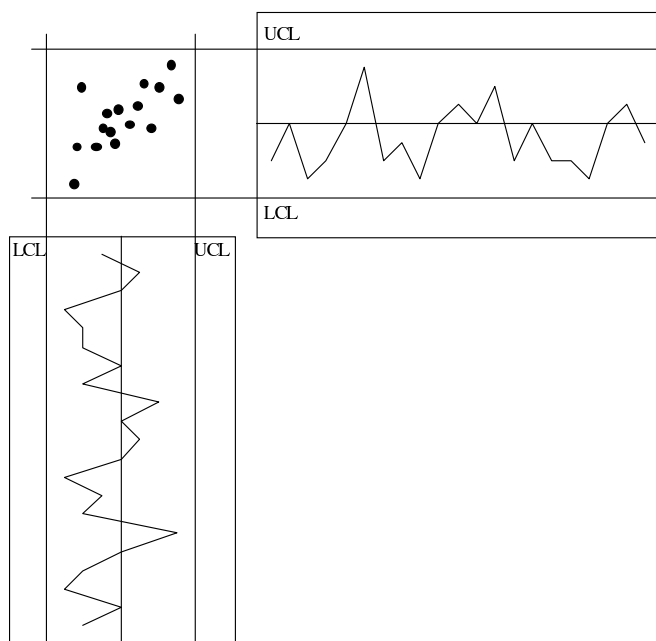
Se fosse stato $S^2 = 14.44$:

$$\frac{(n-1)S^2}{\bar{X}} = \frac{25 \cdot 14.44}{12.3} = 29.35 \in [13.12, 40.65]$$

e in questo caso accetto l'ipotesi di Poissonianità.

13 Carte di controllo multivariate

Fin qui abbiamo considerato l'analisi di una singola caratteristica di qualità. In molti casi, capita di dover analizzare più variabili collegate tra loro. Si può pensare di applicare carte di controllo univariate alle singole variabili, ma questo metodo è inefficiente e può portare a conclusioni errate. Sono necessari metodi multivariati che considerano le variabili congiuntamente. Pensiamo, ad es., ad un anello metallico che ha un diametro interno X_1 e un diametro esterno X_2 . Entrambi determinano l'utilità dell'anello. Supponiamo che X_1 e X_2 siano variabili aleatorie distribuite normali. Si può pensare di fare due carte $\bar{X}$, una per ogni caratteristica. Il processo è considerato in controllo se le medie campionarie $\bar{x}_1$ e $\bar{x}_2$ cadono entro i rispettivi limiti di controllo.



Questo equivale a che la coppia $(\bar{x}_1, \bar{x}_2)$ cada nel quadrato indicato in figura.

In generale avremo un vettore $X = [X_1, X_2, \dots, X_n]$ senza ipotesi di indipendenza e identica distribuzione. Supponiamo, per semplicità, di lavorare senza sottogruppi. L'estensione al caso con sottogruppi è sempre possibile. Si può pensare di usare n carte univariate in parallelo, ognuna con i suoi limiti di controllo. In questo caso, però, abbiamo eccessivi falsi allarmi. Infatti, se per ogni caratteristica $P(\text{falso allarme}) \simeq \alpha$ e le caratteristiche sono indipendenti e identicamente distribuite, risulta

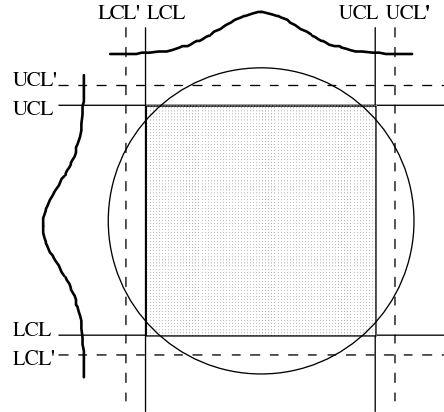
$$P(\text{almeno un falso allarme}) \simeq n\alpha.$$

Ad esempio, per $\alpha = 0.0027$ e $n = 2$, la probabilità di cadere nel quadrato è

$$(1 - \alpha)(1 - \alpha) = 0.9973 \cdot 0.9973 = 0.9946 < 0.9973$$

e quindi la probabilità di falso allarme è maggiore di α .

Un'idea è quella di allargare i limiti di controllo in modo tale che la probabilità di cadere fuori dal quadrato sia α . In questo caso si considerano i limiti alla Bonferroni che consistono nell'allargare i limiti per ogni componente passando da α a α/n (dove n è il numero di elementi del vettore).



$$P(\text{punto fuori dal cerchio}) = \alpha$$

$$P(\text{punto fuori dal quadrato}) = n\alpha$$

$$P(\text{punto fuori dal quadrato tratteggiato}) = \alpha$$

Nel controllo statistico di qualità univariato, generalmente si usa la distribuzione normale per descrivere l'andamento della caratteristica di qualità continua

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad x \in \mathbb{R}.$$

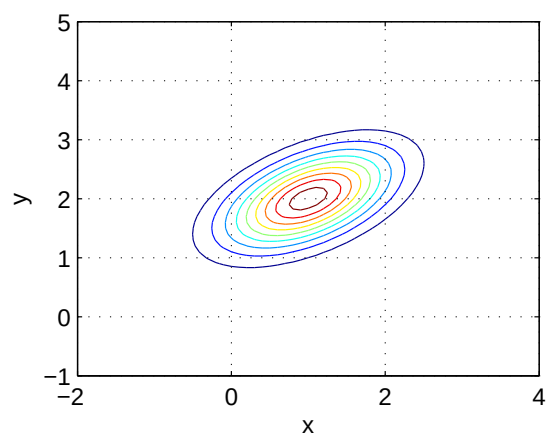
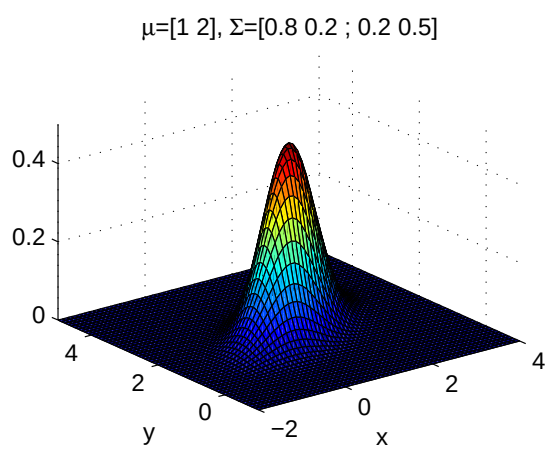
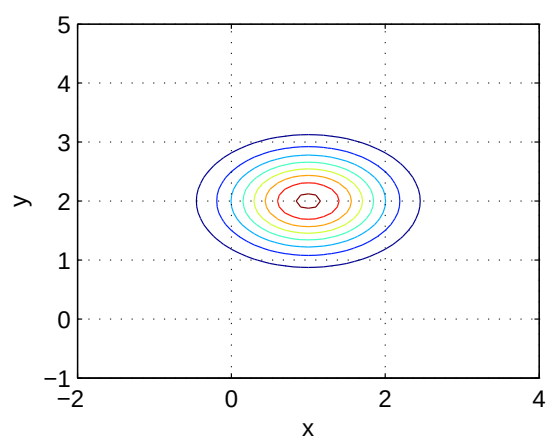
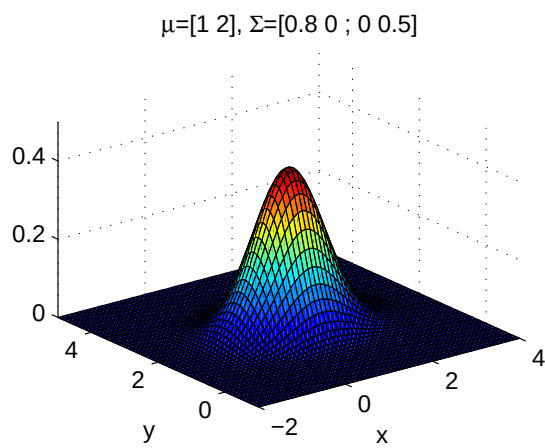
L'analogo nel caso multivariato è la gaussiana multivariata

$$f(X) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} e^{-\frac{1}{2}(X-\mu)'\Sigma^{-1}(X-\mu)} \quad X \in \mathbb{R}^n$$

dove $\mu = [\mu_1, \mu_2, \dots, \mu_n]$ è il vettore delle medie e

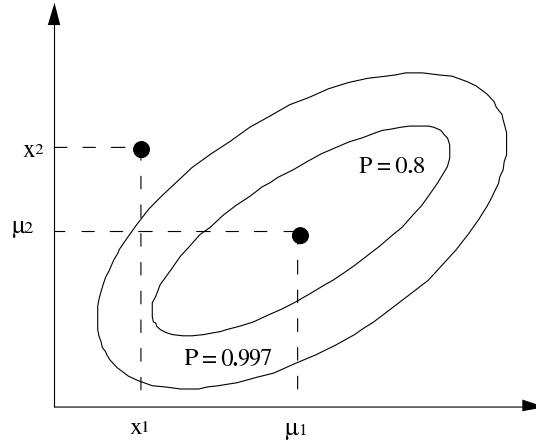
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1,n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2,n} \\ \dots & \dots & \dots & \dots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}$$

Per $n = 2$ si ha la normale bivariata. Le linee di livello della d.d.p. gaussiana bivariata sono delle ellissi con assi paralleli agli assi x e y nel caso in cui le due variabili siano indipendenti o con assi non paralleli a x e y in caso contrario.



Supponiamo che il vettore X abbia distribuzione normale multivariata.

E' giusto allarmarsi quando si osserva un punto su un ellissoide "esterno" (cioè un ellissoide a probabilità bassa).



Osserviamo che, senza indipendenza, la correzione di Bonferroni non è possibile perchè non so calcolare la probabilità globale di falso allarme.

Un punto è tanto più probabile quanto più è vicino al centro dell'ellisse $[\mu_1, \mu_2, \dots, \mu_n]$. Introduco allora una misura di distanza che cresce man mano che X si sposta su ellissi più esterne. Si dice **distanza di Mahalanobis** la distanza

$$Q = (X - \mu)' \Sigma^{-1} (X - \mu)$$

Caso particolare: $n = 2$ e $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$ (cioè due componenti indipendenti). In questo caso

$$Q = \frac{(X_1 - \mu_1)^2}{\sigma_1^2} + \frac{(X_2 - \mu_2)^2}{\sigma_2^2}$$

definisce un'ellisse centrata in $[\mu_1, \mu_2]$ con assi principali paralleli agli assi x_1 e x_2 .

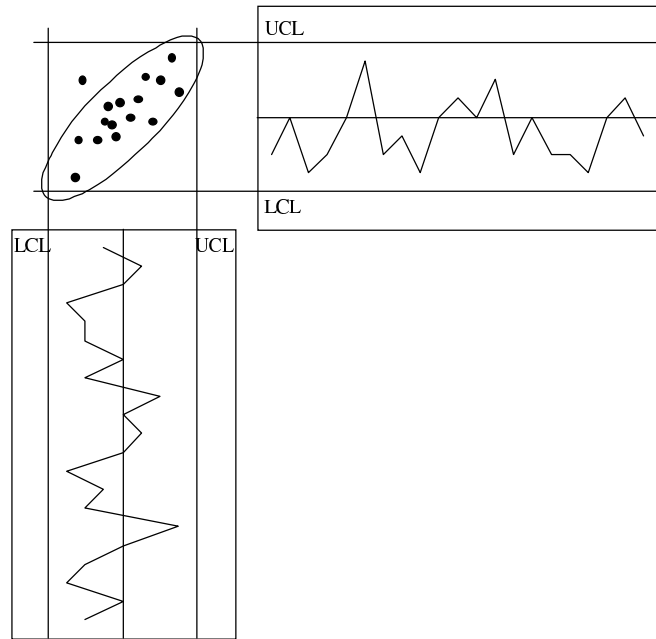
Se $X \sim N(\mu, \Sigma) \Rightarrow Q = (X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi_n^2$.

Se $n = 2$ e fisso α (ad es. $\alpha = 0.0027$), allora trovo $\chi_{\alpha,n}^2$ e traccio l'ellisse di equazione

$$(X - \mu)' \Sigma^{-1} (X - \mu) = \chi_{\alpha,n}^2.$$

Questa è l'**ellisse di controllo**. Segnalo un fuori controllo quando un punto cade fuori da questa ellisse.

Nel caso in cui le due caratteristiche di qualità sono dipendenti ($\sigma_{12} \neq 0$) gli assi principali dell'ellisse non sono più paralleli agli assi x_1 e x_2 .



Dalla figura si nota che il punto che cade fuori dall'ellisse di controllo non avrebbe segnalato un fuori controllo sulle due carte individuali.

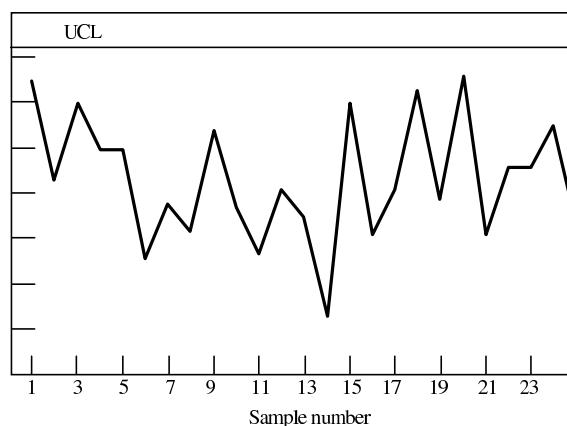
Ci sono due svantaggi associati alle ellissi di controllo:

- si perde la sequenza temporale dei punti;
- difficoltà nel costruire l'ellisse di controllo per $n > 2$.

Per evitare queste difficoltà si calcola Q per ogni campione

$$Q(t) = (X(t) - \mu)' \Sigma^{-1} (X(t) - \mu)$$

e lo si plotta su una carta di controllo detta carta di controllo χ^2 che ha solo un limite superiore $UCL = \chi^2_{\alpha, n}$.



In questo caso si conserva la sequenza temporale. Inoltre, ha il vantaggio che lo “stato” del processo è caratterizzato da un singolo numero (il valore della statistica Q). Questo è particolarmente utile quando ci sono più caratteristiche di interesse.

Nella pratica μ e Σ non sono noti, però si dispone delle stime $\bar{\bar{X}}$ e $\hat{\Sigma}$. Allora si disegnano i punti

$$T^2(t) = (X(t) - \bar{\bar{X}})' \hat{\Sigma}^{-1} (X(t) - \bar{\bar{X}}).$$

In questa forma, la procedura è solitamente chiamata Hotelling T^2 control chart.

Curiosità: il lavoro originario nel controllo di qualità multivariato è dovuto a Hotelling (1947) che ha applicato la procedura ai dati dei dispositivi di puntamento per il lancio delle bombe durante la II guerra mondiale.

Osservazioni per il T^2 senza sottogruppi

- nel caso univariato, T^2 in Fase I coincide con il quadrato di una t di Student;
- se le stime $\bar{\bar{X}}$ e $\hat{\Sigma}$ sono basate su un numero elevato di campioni (almeno 250 se $n \geq 10$ e, in ogni caso, non meno di 100), si pone $UCL = \chi^2_{\alpha, n}$ sia in Fase I che in Fase II;
- per il calcolo di UCL nel caso più generale, non si può usare il χ^2 . Su Montgomery vengono forniti valori di UCL per la Fase I e la Fase II. In particolare, UCL per la Fase I differisce da quello per la Fase II. I dati in Fase I sono utilizzati per vedere se il processo è in controllo e se sono presenti degli outlier da buttare. In questo modo i limiti in Fase II sono più accurati. La differenza tra i limiti di controllo in Fase I e in Fase II, nelle carte univariate, viene di solito ignorata, ma nella carta T^2 è più critica perché ho più parametri da stimare (n parametri per μ e $n(n+1)/2$ per Σ).
Possibile trucco in Fase I: quando calcolo la distanza di Mahalanobis per un certo campione uso media e varianza campionaria stimate usando solo gli altri campioni (“jackknife”), cioè uso i campioni eccetto quello che sto analizzando (se il campione considerato ha varianza molto grande, questo produrrebbe limiti alti e probabilmente non mi risulterebbe il fuori controllo, invece se non lo utilizzo per calcolare media e varianza, allora probabilmente rilevo il fuori controllo) ;
- in Fase I, calcolare $\hat{\Sigma}$ usando varianza e covarianza campionarie non è robusto, allora è meglio stimare Σ usando una specie di Moving Range.

Interpretazione dei fuori controllo

Una difficoltà che si incontra con le carte multivariate è l'interpretazione del fuori controllo. Se la carta T^2 segnala un fuori controllo, quale delle variabili è responsabile (o quali variabili sono responsabili)? Ci sono diversi possibili approcci:

- si disegnano le carte univariate (carte X o $\bar{X}$) con i limiti di controllo alla Bonferroni (questo approccio riduce i falsi allarmi associati con l'uso simultaneo di più carte di controllo univariate). I limiti alla Bonferroni non sono esatti se le variabili sono correlate;
- guardo le carte di controllo come strumenti diagnostici, senza limiti di controllo;

- trasformo le variabili (ad es. calcolo le componenti principali) ed esamino i valori trasformati. In questo caso, però, si perde l'interpretazione fisica;
- decompongo T^2 in modo da evidenziare il contributo delle singole variabili

$$d_i = T^2 - T_{(i)}^2$$

dove T^2 è il valore della statistica, $T_{(i)}^2$ è il valore di T^2 escludendo X_i e d_i è il contributo di X_i al T^2 complessivo. Quando si ha un fuori controllo è bene calcolare i valori dei d_i e porre attenzione alle variabili con d_i grandi.

Violazione delle ipotesi

- *Gaussianità*. La distribuzione di T^2 è sensibile alla skewness (non alla curtosi) e il problema peggiora al crescere di n . Un possibile rimedio è quello di usare una carta MEWMA.
- *Indipendenza*. Nel caso senza sottogruppi, la situazione è simile a quella della carta X , che è ragionevolmente robusta, pur di avere abbastanza campioni. Nel caso multivariato, tuttavia, a causa del numero di parametri da stimare, il numero di campioni necessario potrebbe risultare troppo grande. Come rimedio si può utilizzare una carta multivariata per osservazioni correlate o un run to run control.

Vantaggi della carta T^2

- Metodo di validità generale;
- valutazione rigorosa della probabilità di falso allarme. La valutazione di falso allarme è più rigorosa rispetto all'utilizzo dei limiti di Bonferroni. Infatti, per il calcolo dei limiti di Bonferroni è richiesta l'indipendenza, qui invece no.

Svantaggi della carta T^2

- Ci sono difficoltà ad interpretare i fuori controllo. Come possibile rimedio si possono esaminare le carte univariate;
- ci sono difficoltà a gestire molte variabili (n grande). Un possibile rimedio è la riduzione delle variabili;
- è sensibile all'ipotesi di gaussianità. Come possibile rimedio si può applicare una carta MEWMA;
- sono poco pronte ad identificare piccole derive nella media. Anche in questo caso si può passare ad una carta MEWMA;
- con il T^2 si sintetizza l'informazione in una singola variabile, allora non è detto che tutti i fuori controllo siano rilevabili.

14 Carta MEWMA

Se non ho i sottogruppi, la carta T^2 non è robusta rispetto all'ipotesi di gaussianità. Inoltre, non è sensibile a piccoli spostamenti nella media. Nel caso univariato, per risolvere gli analoghi problemi della carta $\bar{X}$ erano state introdotte le carte EWMA. Qui introduciamo le carte MEWMA (multivariate EWMA = carte multivariate a media mobile esponenziale). Le carte MEWMA sono un'estensione delle carte EWMA. Definiamo la media mobile dei campioni

$$Z(t) = \lambda X(t) + (1 - \lambda)Z(t-1) \quad Z(0) = \mu \quad 0 < \lambda \leq 1.$$

I punti da disegnare sulla carta MEWMA sono

$$T^2(t) = (Z(t) - \mu)' \Sigma_Z^{-1} (Z(t) - \mu)$$

dove

$$\Sigma_Z = \frac{\lambda}{2 - \lambda} \left[1 - (1 - \lambda)^{2i} \right] \Sigma.$$

Quando μ e Σ non sono noti, al loro posto verranno utilizzate le stime $\bar{\bar{X}}$ e $\hat{\Sigma}$.

Quando si ha un fuori controllo per vedere a quale o a quali variabili è dovuto, si possono esaminare le carte EWMA univariate senza limiti di controllo (sola diagnosi) oppure con limiti alla Bonferroni.

Violazione delle ipotesi

- *Gaussianità.* La carta MEWMA è ragionevolmente robusta verso l'ipotesi di gaussianità purché λ sia abbastanza piccolo.
- *Indipendenza.* Se manca l'indipendenza, ci sono problemi nel calcolo dei limiti di controllo. Le cose peggiorano al diminuire di λ . In questo caso si può pensare di passare ad una carta T^2 , oppure di sviluppare una carta multivariata per osservazioni correlate o di eliminare la correlazione ricorrendo al run to run control.

Vantaggi della carta MEWMA

- Meglio della carta T^2 per individuare piccoli spostamenti della media;
- robusta nei confronti dell'ipotesi di gaussianità (se λ non è troppo grande).

Svantaggi della carta MEWMA

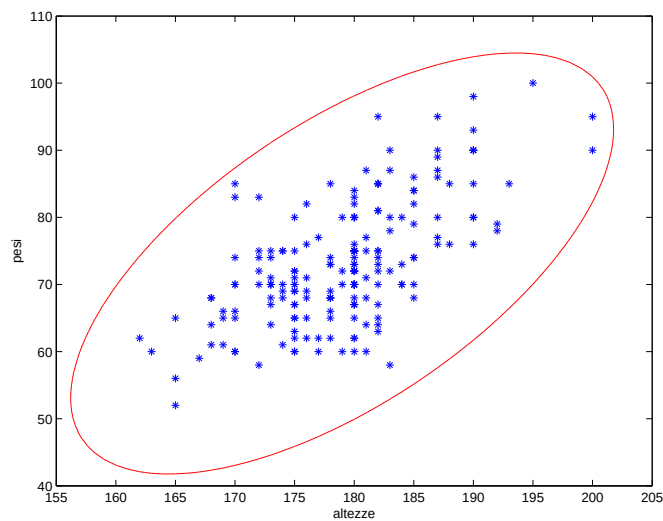
- Non è robusta nei confronti dell'ipotesi di indipendenza e quindi è poco adatta se il processo in controllo ha cicli o trend;
- poco adatta per segnalare discontinuità isolate. Come possibile rimedio si può usare in parallelo una carta T^2 .

Esercizio.

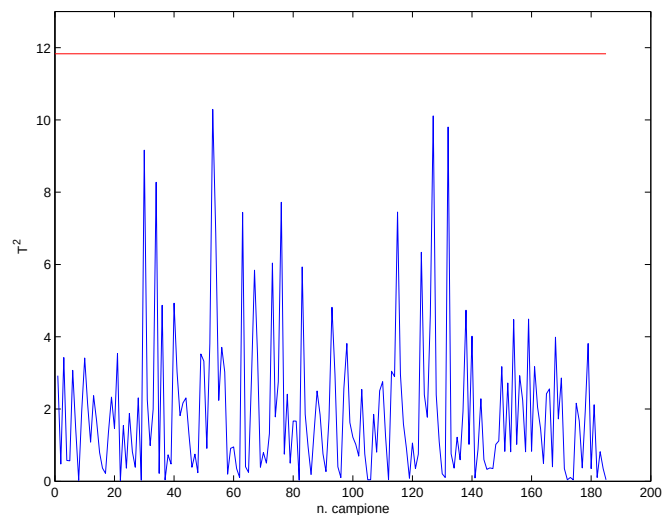
Consideriamo le altezze e i pesi degli studenti maschi del file utilizzato in laboratorio. Vogliamo fare una carta di controllo per la variabile $X = [altezza, peso]$. Supponiamo di avere abbastanza dati per utilizzare l'approssimazione dell'UCL con $\chi^2_{\alpha, n}$. In questo caso $n = 2$ e prendiamo $\alpha = 0.0027$. La media delle altezze è 178.99 cm, mentre la media dei pesi è 73.12 Kg. La matrice di varianza-covarianza è

$$\hat{\Sigma} = \begin{bmatrix} 43.89 & 38.63 \\ 38.63 & 83.19 \end{bmatrix}$$

Andiamo ad analizzare lo scatter plot con l'ellisse di controllo.



Tutti i punti risultano entro l'ellisse di controllo. Se andiamo a disegnare la carta T^2 , possiamo vedere che anche qui non si rilevano fuori controllo ($UCL = 11.829$).

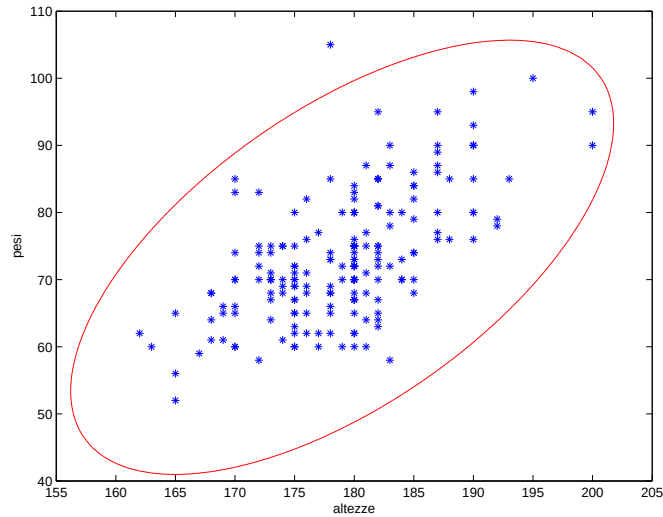


Proviamo, ora a spostare uno dei pesi e a portarlo a 105 kg. La media delle altezze

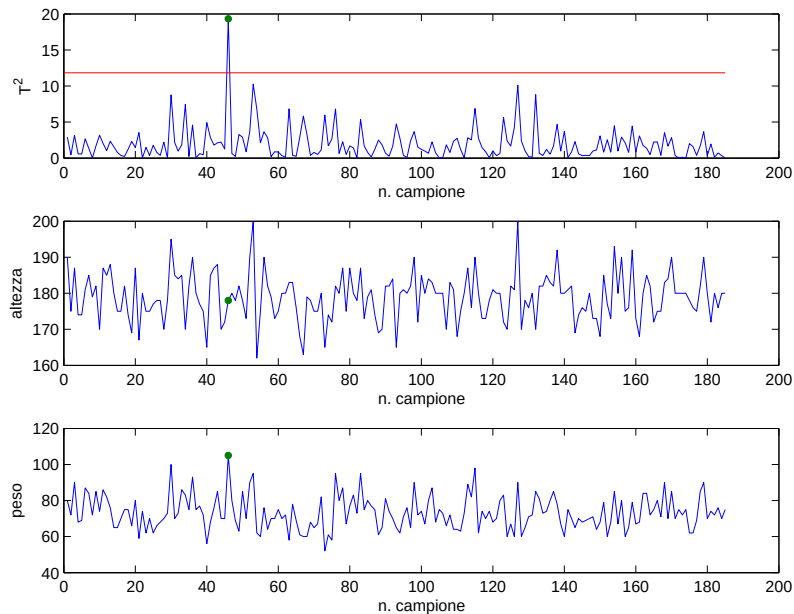
rimane 178.99 cm, mentre la media dei pesi diventa 73.32 Kg. La matrice di varianza-covarianza diventa

$$\hat{\Sigma} = \begin{bmatrix} 43.89 & 38.43 \\ 38.43 & 88.52 \end{bmatrix}$$

Analizziamo lo scatter plot con l'ellisse di controllo



Si vede che un punto è esterno all'ellisse di controllo, ma non si sa a quale delle due caratteristiche è dovuto il fuori controllo e a quale campione è associato. Andiamo allora ad analizzare la carta T^2 insieme alle carte X individuali (senza limiti di controllo in quanto le variabili sono correlate e i limiti alla Bonferroni non sarebbero esatti).



Il fuori controllo si rileva anche dalla carta T^2 ed è relativo al 46-mo campione. Inoltre, dalle carte individuali si vede che il fuori controllo è dovuto alla variabile peso.