

LA STATISTICA DI BASE NELLE APPLICAZIONI INDUSTRIALI

DISTRIBUZIONI.....	2
Raggruppamento in Classi.....	2
ISTOGRAMMA.....	2
FUNZIONE DI (DENSITÀ DI) FREQUENZA.....	2
ISTOGRAMMA CUMULATO.....	3
INDICI DI UNA DISTRIBUZIONE.....	4
Indici di centralità o di posizione.....	4
Indici di variabilità.....	4
LA DISTRIBUZIONE NORMALE.....	6
IL CAMPIONAMENTO.....	7
STIMA DELLA MEDIA E DELLA DISPERSIONE.....	8
Stima degli intervalli di fiducia con grandi campioni ($n > 30$).....	8
Stima degli intervalli di fiducia per piccoli campioni.....	11
Distribuzioni Discrete.....	12
ESERCIZI.....	13
LIMITI DI CONTROLLO.....	14
Limiti estesi.....	16
Cause speciali nelle carte X/R.....	18

DISTRIBUZIONI

Raggruppamento in Classi

Necessario nel continuo
Consente di osservare la distribuzione
Eventuale perdita di precisione e informazione
La scelta delle classi influenza le analisi (comparazioni, χ^2)

Suggerimenti:

quando possibile adottare classi di ampiezza costante
contenere il numero di classi entro **3,3 Log N+1**
più praticamente, anche $\oplus N$ è un buona scelta
indicare le classi con un decimale in più rispetto ai valori della popolazione

ISTOGRAMMA

Rappresentazione grafica della distribuzione in classi dei valori del campione (variabile continua). Ogni classe è rappresentata da un rettangolo, la cui area indica la frequenza del campione in quella classe:

l'area sottesa da una porzione del grafico, determina la frequenza in un determinato campo di variabilità dei valori del campione.

Non è l'altezza dei rettangoli ad indicare la frequenza, ma l'area.

Solo se le classi hanno uguale ampiezza (c), le frequenze (f) sono proporzionali alle altezze (h) dei rettangoli (vedi figura).

L'altezza dei rettangoli è detta **intensità o densità di frequenza** ($h=f/c$)

se le classi hanno diversa ampiezza l'altezza dei rettangoli rappresenta solo la densità di frequenza.

FUNZIONE DI (DENSITÀ DI) FREQUENZA

Aumentando il numero di elementi campionati è possibile aumentare indefinitamente il numero delle classi, riducendone l'ampiezza verso zero e osservare che il profilo dell'istogramma tende a rappresentare una curva continua. In molti casi esiste una funzione matematica $f(x)$ che approssima tale curva. Questa funzione viene detta **funzione di densità di frequenza** o, semplicemente, **funzione di frequenza**.

L'area sottesa dalla funzione di densità, in un determinato intervallo approssima la frequenza del parametro statistico in tale intervallo (integrale della funzione, vedi figura).

Con la funzione matematica di densità si possono calcolare valori teorici di frequenza in intervalli qualsiasi, mentre l'istogramma rappresenta un numero finito di valori effettivamente riscontrati/bili in una popolazione o in un campione, pertanto tale funzione può essere pensata come *modello* della popolazione: il modello sarà tanto più corretto quanto più i valori di frequenza calcolati con l'integrale della funzione si avvicinano a tutti i valori della popolazione nella pratica si fa sempre riferimento alle frequenze relative (percentuali), anziché alle frequenze assolute. In tal caso l'area sottesa dagli istogrammi o dalle funzioni di densità, assomma all'unità.

Anche se il modello di una popolazione è corretto, difficilmente l'istogramma di un piccolo campione assomiglierà alla curva di densità; crescendo la numerosità, però, il campione ha sempre maggiori probabilità di assomigliare alla popolazione da cui è tratto.

La funzione di densità di frequenza relativa è anche detta ***funzione di densità di probabilità*** poiché l'area sottesa dal suo grafico può essere anche intesa quale probabilità di trovare determinati valori in un campione casuale (vedi figura).

ISTOGRAMMA CUMULATO

E' la rappresentazione grafica di una variabile continua *cumulata* per classi. Riducendo progressivamente l'ampiezza delle classi, tendenzialmente a zero, il grafico viene ad approssimare una curva continua che rappresenta una funzione matematica detta *funzione integrale* della funzione di densità.

INDICI DI UNA DISTRIBUZIONE

Osservando la distribuzione di una variabile (continua o discontinua) si può notare che può essere caratterizzata dal modo in cui i valori numerici si accumulano o meno in intervalli specifici e attorno a punti privilegiati del campo di variazione.

Per descrivere una distribuzione si usano fundamentalmente quattro caratteristiche:

1. centralità
2. dispersione
3. simmetria
4. accumulazione

Indici di centralità o di posizione

Il valore più frequente riscontrato (o la classe più frequente) o riscontrabile in base alla funzione di densità, è detto **moda**.

Il valore che divide in due parti il numero dei valori della distribuzione è detto **mediana**. Tale valore si determina dall'istogramma cumulato in corrispondenza della frequenza relativa 0,5 (frequenza percentuale 50%).

(Analogamente si parla di *percentili*, *decili*, *quartili*, etc. intendendo i limiti delle classi che contengono, ognuna, 1%, 10%, 25%, etc dei valori riscontrati).

Il valore che "equilibra" l'asse delle ascisse su cui "cadono" i valori è detto **media aritmetica** ($\sum x_i / N$).

Si usano a volte altri tipi di medie (geometrica, armonica, etc.) ma di norma si usa la media aritmetica, denominandola semplicemente **media**.

Nessuno di questi indicatori (statistiche campionarie) è sufficiente da solo a descrivere compiutamente una distribuzione:

- la moda può rappresentare più punti di accumulazione (distribuzioni bimodali) e spesso dipende dal raggruppamento dei dati
- la mediana non è facilmente manipolabile matematicamente
- la media spesso non indica bene dove i valori si concentrano; la media è però un operatore matematico e si presta bene ad entrare nel calcolo

Indici di variabilità

Attorno ai punti centrali della distribuzione si tende ad avere la massima concentrazione dei valori; ma come e quanto si "disperdono" i valori attorno a questi punti? Si utilizzano varie *statistiche di dispersione*:

l'escursione o range, differenza tra il valore massimo e quello minimo: è significativo solo quando si vuol essere prudenti! E poi oscilla sempre più (anziché convergere) al crescere della numerosità del campione.

lo *scarto assoluto medio dalla media* $\Sigma|x_i-\mu|/N$ con il quale si hanno, però, difficoltà matematiche (es.:derivazione dei valori assoluti)

lo *scarto quadratico medio dalla media* $\Sigma(x_i-\mu)^2/N$ sul quale è facile operare matematicamente. Con questa dizione si preferisce, però, riferirsi alla radice quadrata della media dei quadrati degli scarti dalla media: $\sigma=\sqrt{(\Sigma(x_i-\mu)^2/N)}$

Per evitare confusioni, nella pratica industriale σ viene denominato *deviazione standard* (USA) e *scarto tipo* (UNI)

Il quadrato di σ è denominato *varianza*.

LA DISTRIBUZIONE NORMALE

Gran parte dei fenomeni naturali, industriali, economici, etc., generano distribuzioni descritte da un modello particolare, rappresentato matematicamente dalla *funzione di Gauss o gaussiana*. Questa funzione ha il noto andamento a campana riportato in figura.

Notiamo che la curva è simmetrica, distribuita uniformemente attorno ad un unico valore centrale che coincide con la moda, con la mediana e con la media aritmetica. $f(x_0)$ rappresenta la densità di probabilità in un intorno infinitesimale di x_0 .

La funzione integrale, calcolata nell'intervallo $x_0 \text{ -- } x_1$ fornisce la probabilità di ottenere, estraendo a caso un elemento da una popolazione che segue il modello normale, un valore compreso in tale intervallo.

I valori della funzione integrale sono tabulati e quindi facilmente accessibili senza calcolo, per il caso in cui la funzione di probabilità abbia le seguenti caratteristiche:
Media=0

Deviazione standard=1

(Area sottesa=1, ma questo è vero per tutte le funzioni di probabilità)

La gaussiana con queste caratteristiche si dice *standardizzata* ed è descritta dall'equazione

$$p(z) = (1/\sqrt{2\pi})e^{-(z^2)/2}$$

L'unità di misura sulle ascisse della funzione di probabilità standardizzata diventa pari a σ , la deviazione standard.

Qualunque gaussiana può essere standardizzata, con una trasformazione degli assi cartesiani equivalenti ad un cambio di unità di misura e ad uno spostamento dell'origine sulle ascisse.

Per conoscere il valore dell'integrale di una qualunque gaussiana (ovvero la probabilità in un certo intervallo), basta cercare nella tabella degli integrali della gaussiana standardizzata in corrispondenza di $z=(x-\mu)/\sigma$

Esaminando la tabella degli integrali della gaussiana notiamo che:

nell'intervallo $\pm 3\sigma$ cadono il 99,73% dei valori (abbiamo il 99,73% di probabilità di estrarre un valore dell'intervallo)

nell'intervallo $\pm 2\sigma$ cadono il 95,45% dei valori

nell'intervallo $\pm 1\sigma$ cadono il 68,27% dei valori

DEFINIZIONI

Parametri: caratteristiche della popolazione (universo)

Statistiche: caratteristiche di un campione

IL CAMPIONAMENTO

Scopo principale della statistica è determinare (inferire) le *probabili* caratteristiche della popolazione o, viceversa, dedurre dai parametri della popolazione le caratteristiche *probabili* di un campione.

Esempi:

Stima della qualità di una produzione dall'esame di alcuni campioni

Stima del numero di difetti in una partita proveniente da un lotto prodotto con AQL=0,5%

Ma con quale *fiducia* accettiamo, ad esempio, per la popolazione le statistiche rilevate in un campione? Grazie alle proprietà scoperte nella distribuzione normale, siamo spesso in grado di calcolare la probabilità degli scostamenti delle statistiche rispetto ai parametri.

La **legge dei grandi numeri** ci dice che se si estraggono ripetuti campioni di dimensione n da una popolazione qualunque (anche non "normale") che abbia media μ e varianza σ^2 , all'aumentare di n la distribuzione campionaria delle medie dei campioni tenderà ad avvicinarsi alla normalità e avrà come media μ e come varianza σ^2/n (la distribuzione della media delle medie di campioni molto grandi è *normale*).

Pertanto, la media M di una caratteristica rilevata in un campione, ha probabilità definite (dalla tabella della gaussiana standard) di scostarsi dalla media μ della popolazione; per es. la probabilità che tale scostamento sia compreso nell'intervallo $\pm 2\sigma/\sqrt{n}$ è del 95,45%. In altre parole possiamo dire di avere un **grado di fiducia** del 95,45% di aver trovato la media della popolazione in detto intervallo detto **intervallo di fiducia**.

In generale si può dire che la media di una popolazione è stimata con:

$$\mu = M \pm z\sigma/\sqrt{n}$$

Z è detto **coefficiente di fiducia** (*confidenza*) e determina l'**intervallo di fiducia**

E' d'uso chiedere un grado di fiducia del 95% al quale corrisponde un coefficiente $z=1,96$ (vedi tabella)

STIMA DELLA MEDIA E DELLA DISPERSIONE

Se m e s rappresentano la media e lo scarto quadratico medio¹ di un campione casuale di dimensione n , possiamo **stimare** la media e lo scarto tipo della relativa popolazione come segue:

$$\mu \cong m$$
$$\sigma \cong s \cdot \sqrt{\frac{n}{n-1}} = s'$$

Queste stime sono tanto migliori quanto più numeroso è il campione; gli errori di stima si riducono al crescere della numerosità n del campione, ma la statistica ci consente di valutarli in ogni caso (utilizzando le distribuzioni matematiche t e χ^2).

Stima degli intervalli di fiducia con grandi campioni ($n > 30$)

Dato un intervallo di valori attorno alla stima della media o dello scarto tipo, possiamo calcolare la probabilità con cui il parametro della popolazione cade in tale intervallo (detto *intervallo di fiducia*).

Per valutare questo intervallo, nel caso di stima della media della popolazione, consideriamo che la media di un qualunque campione è uno dei valori della *distribuzione della media campionaria*, la cui media coincide con quella della popolazione; questa distribuzione è normale quando n è grande (vedi teorema dei grandi numeri) e pertanto di essa sappiamo calcolare la probabilità (o percentuale) con cui può apparire ogni suo valore, ovvero la media di ogni possibile campione. In altre parole, conoscendo lo scarto tipo della distribuzione della media campionaria (il teorema dei grandi numeri ci dice che vale σ/\sqrt{n})² siamo in grado di valutare la

¹ Lo scarto quadratico medio di un campione è $s = \sqrt{\frac{\sum (x_i - m)^2}{n}}$. Se anziché dividere per n si

divide per $n-1$ si ottiene uno stimatore dello scarto quadratico medio della popolazione che indichiamo con s' . Tale stimatore si ottiene anche moltiplicando lo scarto quadratico medio del

campione per il fattore $\sqrt{\frac{n}{n-1}}$; risulta infatti: $s' = \sqrt{\frac{\sum (x_i - m)^2}{n-1}} = s \cdot \sqrt{\frac{n}{n-1}}$

probabilità con cui può apparire un suo qualunque valore - rappresentante la media di un campione - compreso in un dato intervallo ed in particolare nell'intervallo $|m-\mu|$. Sappiamo che questa probabilità corrisponde all'area sottesa dalla curva normale in quell'intervallo, misurato in unità pari allo scarto tipo della curva stessa. Nel caso della distribuzione della media campionaria l'intervallo ha ampiezza:

$$z_c = |\mu - m| / (\sigma / \sqrt{n})$$

Con tale valore possiamo accedere alle tabelle che forniscono la probabilità con cui si può attendere un valore compreso tra la media campionaria - coincidente con la media della popolazione - e la media rilevata nel campione; indicando con $P(z_c)$ tale probabilità, potremmo quindi dire che **la media della popolazione si discosta, con probabilità $P(z_c)$, dalla media del campione di al più $z_c(\sigma/\sqrt{n})$.**

L'intervallo $|\mu - m|$ è l'*intervallo di fiducia* e $P(z_c)$ è il *livello o grado di fiducia*. I valori

$$m \pm z_c(\sigma/\sqrt{n})$$

sono detti *limiti di fiducia o di confidenza*³.

Piuttosto che calcolare il livello di fiducia da assegnare ad un dato intervallo, spesso si cerca di determinare questo intervallo in funzione del livello di fiducia richiesto. Le tavole della distribuzione normale vengono allora impiegate all'opposto, ricavando il valore z_c . Nella pratica statistica sono richiesti frequentemente i seguenti livelli di fiducia, a cui corrispondono valori tipici di z_c indicati:

z_c	Livello di fiducia
0.67	50%
1,28	80%
1,64	90%
1.96	95%
2,33	98%
2,58	99%

L'intervallo corrispondente al livello di fiducia del 50% è detto *errore probabile*.

L'errore probabile non deve essere confuso con l'*errore standard*, con il quale si intende - per qualunque statistica campionaria S - proprio lo scarto quadratico medio

² Ciò vale nell'ipotesi di popolazione infinita o anche finita se il campione è scelto con ripetizione. In pratica si accetta questa ipotesi quando $n \gg N$; altrimenti lo scarto tipo della media campionaria è determinato da $\sigma/\sqrt{n} \cdot \sqrt{[(N-n)/(N-1)]}$

³ Nel caso non si possa accettare l'ipotesi di popolazione finita senza ripetizione i limiti di fiducia diventano:

$$m \pm z_c(\sigma/\sqrt{n}) \cdot \sqrt{[(N-n)/(N-1)]}$$

della distribuzione (deviazione standard). L'errore standard della distribuzione dello scarto tipo campionario risulta: $\sigma_s = \frac{\sigma}{\sqrt{2N}}$ e pertanto l'intervallo di fiducia dello scarto tipo della popolazione può essere calcolato con:

$$\sigma = s \pm z_c \sigma_s$$

ovvero:

$$\sigma = s \pm z_c \frac{\sigma}{\sqrt{2N}}$$

quindi avremo: $\sigma\sqrt{(2N)}=s\sqrt{(2N)}\pm z_c\sigma$ da cui possiamo ricavare per la popolazione:

$$\sigma=s\sqrt{(2N)}/(\sqrt{(2N)}\pm z_c)$$

Quando σ non è conosciuto, viene stimato con lo stesso s o, meglio, con s' .

Esercizio:

50 alberini prelevati a caso da un lotto di 2000, hanno dimostrato un diametro medio di 17,3 mm con scarto tipo di 0,20 mm. Valutiamo la media dei diametri dell'intero lotto: assumendo quale media di tutti gli alberini la media rilevata nel campione commetteremo un errore probabile pari a

*$\pm 0,67s'/\sqrt{N} = 0,67*0,20*\sqrt{(50/49)}/\sqrt{50} = \pm 0,019$ (la media si troverebbe in questo intervallo con una probabilità del 50%).*

*Con probabilità del 95%, invece, la media della popolazione dista da 17,3 per meno di $1,96*0,20/\sqrt{49}=0,056$*

In altre parole possiamo ritenere, con un grado di fiducia del 95%, che la media di tutti gli alberini sia compresa tra $17,3\pm 0,056$, cioè tra 17,24 e 17,36.

*Abbiamo così stimato lo scarto tipo della popolazione con $0,2\sqrt{(50/49)}$. Assumendo come scarto tipo della popolazione quello del campione, si commetterebbe un errore probabile pari a $\pm 0,67*0,2\sqrt{(50/40)}/\sqrt{(2*50)} = \pm 0,01498$.*

Lo scarto tipo effettivo della popolazione può essere determinato entro un intervallo di confidenza con la formula sopraccitata. Per esempio, potremmo dire che sarà compreso nell'intervallo di fiducia del 95% pari a:

*$0,2\sqrt{(2*50)}/(\sqrt{(2*50)}\pm 1,96)$, ovvero tra 0,16722 e 0,248756.*

Più semplicemente potremmo approssimare lo stesso intervallo come :

$$0,2\pm 1,96*0,2/\sqrt{(2*49)},$$

I limiti di fiducia diventerebbero allora: 0,16 e 0,24, che confermiamo essere una buona approssimazione di quelli sopra determinati.

Stima degli intervalli di fiducia per piccoli campioni

Quando $n < 30$ la stima della media e dello scarto tipo non può effettuarsi impiegando la distribuzione normale. Per la media si impiega la distribuzione t , detta anche di Student e per lo scarto tipo la distribuzione χ^2 .

Queste distribuzioni vengono approssimate sempre meglio dalla distribuzione normale quando $n > 30$ e potrebbero essere usate in ogni caso, per qualunque numerosità dei campioni. Esse sono, però, di uso più complesso e pertanto impiegate solo con i piccoli campioni che si richiede siano estratti da una popolazione normale (condizione non necessaria con i grandi campioni).

Le due distribuzioni sono definite dalle seguenti statistiche:

$$t = (m - \mu) / (s' / \sqrt{n}) = (m - \mu) \cdot (\sqrt{(n-1)} / s)$$

$$\chi^2 = n \cdot s^2 / \sigma^2$$

entrambe tabulate per i valori notevoli di uso pratico.

I limiti di confidenza per la media diventano:

$$\boxed{m \pm t_c \cdot s' / \sqrt{n} \quad \text{oppure} \quad m \pm t_c \cdot s / \sqrt{(n-1)}}$$

Per quanto riguarda lo scarto tipo occorre precisare che la distribuzione χ^2 non è simmetrica e pertanto ad ogni livello di confidenza corrispondono due diversi valori dell'area sottesa a destra e a sinistra dell'intervallo di fiducia.

I limiti per lo scarto tipo sono quindi indicati da :

$$\boxed{s \cdot \sqrt{n / \chi_{1-\alpha}} < \sigma < s \cdot \sqrt{n / \chi_{\alpha}}}$$

dove α indica l'area sottesa a destra del limite superiore.

Parte 3a

Distribuzioni Discrete

Quando l'esame di una caratteristica discreta fornisce un dato risultato con probabilità fissa p , diventa utile l'impiego della **distribuzione binomiale** (o di Bernoulli). Questa è una distribuzione di probabilità che indica la probabilità $P(k)$ di individuare, in un campione casuale di dimensione n , k elementi con la caratteristica detta⁴:

$$P(k) = \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

dove $q=1-p$ è la probabilità che la caratteristica *non* abbia il valore cercato. Nella gestione della qualità tale caratteristica è tipicamente la conformità ad un requisito specificato.

L'impiego di questa distribuzione ha un limite pratico nella complessità del calcolo. Fortunatamente la distribuzione binomiale è sufficientemente ben approssimata da altre funzioni matematica di impiego più semplice, tra le quali la distribuzione normale e la **distribuzione di Poisson**:

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

dove $\lambda = np$. Questa approssimazione risulta valida nel caso di *evento raro*, cioè quando n è grande, p è quasi zero e q è prossimo a 1. Infatti la funzione di Poisson è il limite di quella di Bernoulli per $n \rightarrow \infty, p \rightarrow 0$.

In pratica si accetta l'approssimazione di Poisson quando $n > 50$ e $np < 5$.

Se n è grande, ma né p né q sono prossimi allo zero, la binomiale viene bene approssimata dalla distribuzione di Gauss che conosciamo.

In pratica l'approssimazione della curva normale è accettabile quando sia np che nq risultano maggiori di 5.

La seguente tabella indica le formule di calcolo della media e dello scarto tipo per le distribuzioni di Bernoulli e di Poisson:

<i>Statistica</i>	<i>Distribuzione Binomiale</i>	<i>Distribuzione di Poisson</i>
Media	$\mu = np$	$\mu = \lambda$
Scarto Quadratico Medio	$\sigma = \sqrt{npq}$	$\sigma = \sqrt{\lambda}$

⁴ Per il calcolo dei fattoriali ricordare che $x! = x \cdot (x-1) \cdot (x-2) \cdot \dots \cdot 3 \cdot 2 \cdot 1$

Notare, inoltre, che $1! = 0! = 1$

ESERCIZI

1. Un fornitore dichiara un processo con difettosità inferiore al 2%. Qual è la probabilità di trovare al più 1 pezzo difettoso in un lotto di 80 prodotti?
2. Nel lotto precedente vengono rilevati ben 5 pezzi difettosi. Che probabilità abbiamo che il processo sia regolato proprio come dice il fornitore?
3. Il diametro medio interno di un campione di 200 rondelle risulta 0,502 cm, mentre lo scarto quadratico medio è 0,005 cm. I limiti di tolleranza a disegno, per il diametro interno, sono 0,496 e 0,508 . Determinate la percentuale di rondelle difettose che il processo produce
4. Un dado viene lanciato 120 volte. Quante volte devo aspettarmi che si presenti la faccia 4? Che probabilità ho che la faccia 3 si presenti 18 volte o meno?
5. Il 10% degli alberini prodotti è difettoso. Quale probabilità si ha che un campione casuale di 400 alberini mostri al massimo 30 difettosi? Qual è la probabilità di trovare tra 30 e 50 alberini difettosi?
6. In un campione 30 pezzi si riscontra una media di 67,45 con scarto tipo 2,92. Determinare la media e lo scarto tipo della popolazione.
7. Calcolare i limiti di fiducia al 80% e al 95% per la media della popolazione suddetta.
8. Le tensioni di 50 batterie dello stesso tipo hanno una media di 18,2 V ed uno scarto tipo medio di 0,5 V. Trovate l'errore probabile della media e i limiti di fiducia al 50%
9. Una misura è stata registrata come 216,480 g, con un errore probabile di 0,272 g. Quali sono i limiti di fiducia al 95% della misura.
10. Lo scarto quadratico medio dei tempi di vita di un campione di 200 lampadine è stato calcolato nella misura di 100 ore. Trovate i limiti di fiducia al 95% per lo scarto quadratico medio di tutte le lampadine.
11. Quanto grande vorremmo il campione precedente per essere fiduciosi al 99,73% che la deviazione standard della popolazione non differisca dalla deviazione standard campionaria per più del 5%?
12. La tensione media delle batterie prodotte da una fabbrica è 45,1 Volt e lo scarto tipo medio è di 0,04 Volt. Trovare i limiti di fiducia al 95% della tensione di un gruppo di 4 batterie collegate in serie.

LIMITI DI CONTROLLO

I limiti di controllo di una qualunque statistica S a distribuzione normale, che si voglia controllare con un livello di fiducia del 99.73%, possono essere definiti da:

$$S_{\text{sup}} = \bar{S} \pm 3\sigma_S$$

dove σ_S viene determinato in funzione della deviazione standard della popolazione⁵.

Nell'uso con le carte di controllo di Schewhart, però, il calcolo dei limiti prescinde dal tipo di distribuzione, pur usando la formula sopraccitata, poiché questa comunque risulta ben adeguata ad individuare un limite oltre il quale conviene considerare il processo "fuori controllo".

Nella seguente figura si vede che le distribuzioni usuali non normali, che si incontrano nella pratica industriale, forniscono tutte oltre il 98% dei valori nell'intervallo $\pm 3\sigma$.

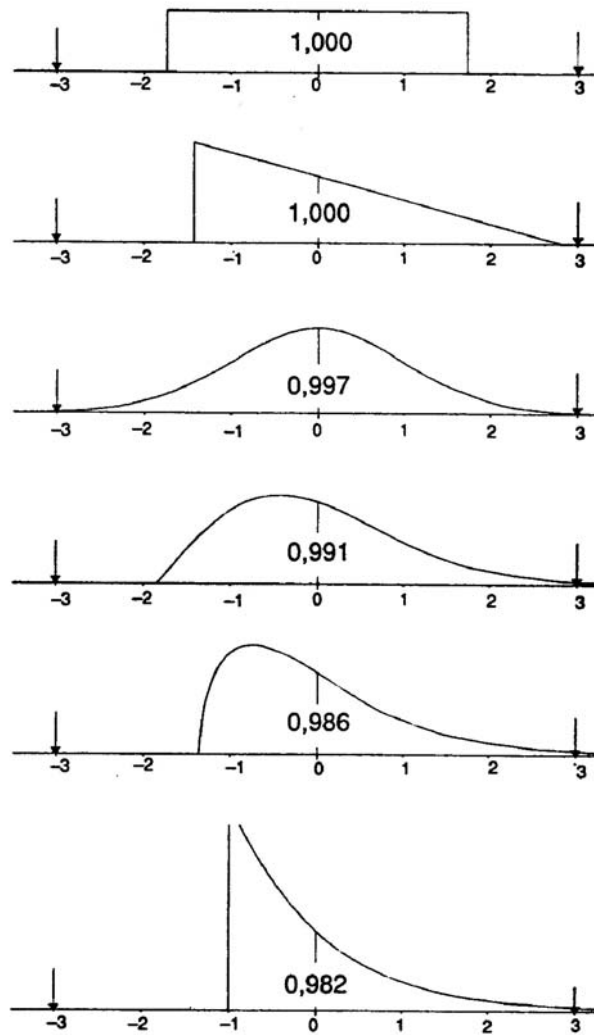
Più in generale Shewhart aveva a suo tempo dimostrato che per le distribuzioni non normali valgono i seguenti criteri:

1. nell'intervallo ± 1 sigma attorno alla media sono compresi all'incirca dal 60% al 75% dei dati
2. nell'intervallo ± 2 sigma attorno alla media sono compresi all'incirca dal 90% al 98% dei dati
3. nell'intervallo ± 3 sigma attorno alla media sono compresi all'incirca dal 99% al 100% dei dati

E' sulla base di questi criteri (e non altro) che si definiscono i limiti di controllo moltiplicando per 3 lo scarto tipo della statistica in oggetto (la media o l'escursione dei campioni, nelle carte X/R).

⁵ Quando S è la media campionaria \bar{X} , per esempio, sappiamo che $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$; per la deviazione

standard campionaria sussiste la relazione $\sigma_S = \frac{\sigma}{\sqrt{2n}}$.



Anche nel calcolo dello scarto tipo, Shewhart pare fare riferimento ad una popolazione con distribuzione normale, sebbene questa non sia mai riscontrabile in pratica. E' stato, infatti, dimostrato che l'assimilare una qualunque distribuzione al modello normale comporta solamente un aumento trascurabile dei segnali "fuori limite": con riferimento alle distribuzioni standardizzate non normali sopra raffigurate, i segnali extra raggiungono il 2% solo nel caso delle ultime due distribuzioni (chi-quadro ed esponenziale).

Riferendoci alla carta X,R i limiti di controllo vengono calcolati a partire dalle seguenti relazioni, stabilite dall'analisi statistica, tramite coefficienti tabellati in funzione del numero e della numerosità dei campioni:

$$\sigma_R = d_3 \sigma \quad \sigma = \frac{\bar{R}}{d_2}$$

per cui abbiamo:

$$UCL_{\bar{R}} = \bar{R} + 3\sigma_R = \bar{R} + 3d_3\sigma = \bar{R} + 3d_3 \frac{\bar{R}}{d_2} = \bar{R} \left(1 + \frac{3d_3}{d_2}\right) = D_4 \bar{R}$$

analogamente definiamo:

$$LCL_{\bar{R}} = D_3 \bar{R}$$

$$UCL_{\bar{X}} = \bar{\bar{X}} + 3 \frac{\bar{R}}{d_2 \sqrt{n}} = \bar{\bar{X}} + A_2 \bar{R}$$

$$LCL_{\bar{X}} = \bar{\bar{X}} - A_2 \bar{R}$$

Ripetiamo che l'impiego di $\pm 3\sigma$ vuole solo applicare il criterio n°3 sopraccitato: nulla ha a che vedere con il controllo del 99,73% dei valori di una gaussiana, poiché queste formule prescindono dalla forma di tutte le distribuzioni in gioco (popolazione di base, medie ed escursioni). Si noti anche, che mentre la distribuzione delle medie potrebbe essere assimilata ad una gaussiana, ciò non sarebbe comunque possibile per la distribuzione dei range.

Limiti estesi

In presenza di una "causa speciale" la dispersione della popolazione aumenta e diventa:

$$\sigma = \sqrt{\left[\frac{\bar{R}}{d_2}\right]^2 + \sigma_{SP}^2}$$

la dispersione aggiuntiva "muove" le medie dei campioni⁶, che ora si discostano dalla distribuzione campionaria che si avrebbe nel processo stabile. I limiti di controllo vengono quindi superati ad indicare la presenza della perturbazione⁷.

⁶ Potrebbe muovere la media dei range, ma questo è il caso meno comune; il ragionamento seguente non cambierebbe.

⁷ Aumentando la numerosità del campione aumenta la sensibilità della carta a rilevare eventuali movimenti delle medie (i limiti si restringono), ma si rischia di avere campioni non più rappresentativi della variabilità intrinseca del processo.

Quando tale perturbazione non possa essere convenientemente eliminata e non impedisca le possibilità di mantenere comunque il processo entro le tolleranze (per esempio nel caso di oscillazioni contenute), occorre individuare diversamente i limiti di controllo, non più calcolabili solo con la media delle escursioni, che rappresenta solo la variabilità dovuta alla cause comuni. In questi casi si parla di *limiti estesi*.

Per il calcolo dell'estensione dei limiti si utilizzano diversi metodi:

- stima dell'intervallo di oscillazione delle medie, da aggiungere ai limiti del processo soggetto alle sole cause comuni
- sostituzione dei limiti "regolari" con un intervallo pari ad un multiplo della deviazione standard delle medie dei campioni; in pratica si usa questo metodo sottintendendo che la distribuzione delle medie dei campioni risulti normale. In tal caso, l'esperienza suggerisce di limitare l'intervallo a ± 2 volte la dispersione.
- sostituzione dei limiti regolari con un intervallo pari a $\pm 3S/\sqrt{n}$ dove S è lo scarto tipo complessivo calcolato su tutti i valori misurati. Anche in questo caso vari autori suggeriscono di verificare la normalità della distribuzione delle medie, mentre, alla luce di quanto sopraddetto, ciò non appare essenziale: le carte di controllo funzionano bene anche senza ipotesi di normalità.
- Allargamento dei limiti naturali con un multiplo (generalmente 1,5) dello scarto tipo delle medie dei campioni calcolato con ANOVA.

In pratica si deve ricorrere ai limiti estesi quando le medie dei campioni mostrano una loro propria variabilità, indipendente dalla variabilità stocastica del processo, o, in altre parole, quando la variabilità *nei* campioni è significativamente inferiore alla variabilità *tra* i campioni⁸.

⁸ Il test F consente di stabilire il caso

CAUSE COMUNI E CAUSE SPECIALI

Cause speciali: di norma sono sufficienti azioni LOCALI

Cause comuni: di norma richiedono interventi sul SISTEMA (85% dei casi)

La confusione è all'origine di molti costi:

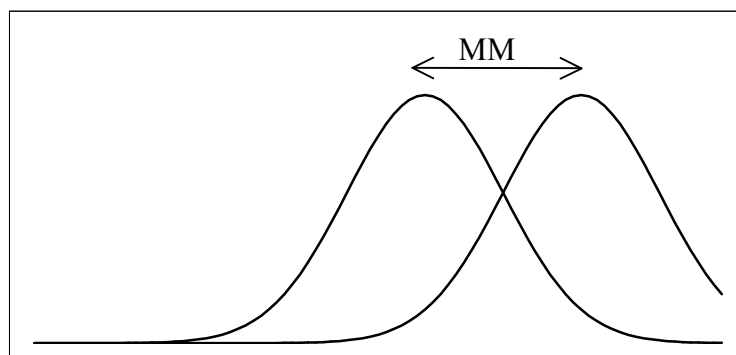
- sforzi inutili
- ritardi di risoluzione dei problemi
- aggravamento dei problemi (es. overadjustment)

Es. I problemi che insorgono a causa di disomogeneità nei materiali di approvvigionamento: richiedono interventi sul sistema, è dannoso intervenire sul processo

Cause speciali nelle carte X/R

Pattern non randomici indicano generalmente la presenza di fenomeni fuori controllo. Ciò non è sempre vero. Una causa speciale apparente nelle carte di controllo è, per esempio, quella dovuta alla variabilità del ritmo di produzione o alla variabilità della frequenza di prelievo. Queste cause apparenti devono essere evitate perché interferiscono nell'analisi della carta di controllo, in particolare nell'analisi delle carte a media variabile (mobile).

Un'altra causa speciale di comune riscontro è l'usura degli utensili. Questo caso può essere assimilato ad un processo con distribuzione statistica mobile o, più precisamente, ad un processo a media mobile.

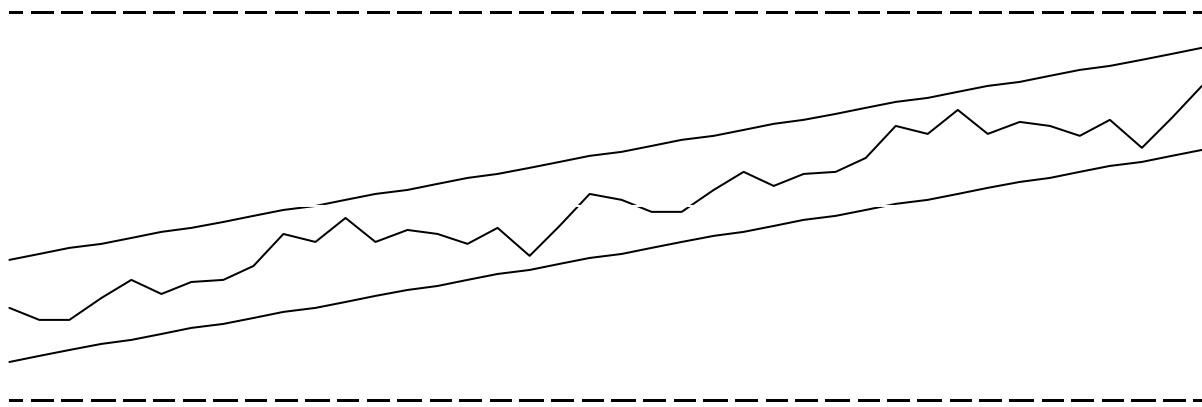


In tal caso si procede a studiare la variabilità delle medie dei campioni x_m , non più rispetto alla loro media X_m , ma rispetto alla linea di interpolazione che rappresenta il trend associato all'usura. Rispetto a questa linea si disegnano quindi i limiti di controllo $\pm A_2 R_m$

I limiti di controllo complessivi del processo, quelli che tengono conto anche del movimento della media, possono servire ad indicare il momento in cui bisogna intervenire sul processo per sostituire l'utensile usurato.

Tali limiti si ottengono semplicemente "allargando" il campo di controllo per un'estensione pari a quella del movimento medio della media x_m osservata su più cicli di usura.

Indicando il movimento della media con MM (max x_m osservato – min x_m osservato) i limiti complessivi di controllo diventano: $X_m \pm 0,5 MM \pm A_2 R_m$



Nel calcolo del Cp e del Cpk, occorre usare qualche accorgimento, affinché gli indici risultino "penalizzati" dal movimento della media.

Tenendo presente che σ rappresenta la variazione dovuta alle cause comuni (variazione determinabile dalla carta R), possiamo usare due diverse correzioni nel calcolo degli indici di capacità:

<p>a) sottrarre al numeratore una quantità pari al movimento della media (ovvero considerare un intervallo di tolleranza (Ls-LI) ridotto)</p>	<p>$C_p = (LS - LI - MM) / 6\sigma$ $C_{pk} = \min[(LS - X_m - MM/2) / 3\sigma; (X_m - LI - MM/2) / 3\sigma]$</p>
<p>b) aumentare il denominatore di una quantità pari al movimento della media</p>	<p>$C_p = (LS - LI) / (6\sigma + MM)$</p>

(ovvero considerare il movimento della media come parte della tolleranza naturale del processo)

$$Cpk = \min\left[\frac{LS - X_m}{3\sigma + MM/2}; \frac{X_m - LI}{3\sigma + MM/2}\right]$$