

GLI STRUMENTI PER LA GESTIONE TOTALE DELLA QUALITÀ

Materiale tratto dal testo "Total Quality Management" – Andrea Chiarini, edito da Franco Angeli. Ogni riproduzione anche parziale è vietata

1. Generalità

I modelli TQM si basano sull'uso di strumenti, sia manageriali che operativi, a tutti i livelli.

I paragrafi successivi trattano gli strumenti maggiormente utilizzati dalle organizzazioni, in particolare:

- foglio raccolta dati;
- stratificazione;
- diagramma causa - effetto;
- istogramma;
- diagramma di Pareto;
- analisi di correlazione;
- carte di controllo.

Queste ultime saranno analizzate nell'ambito del capitolo 9 dedicato al cosiddetto *SPC (Statistical Process Control)*.

Gli strumenti sopra citati sono anche riportati in letteratura come "i sette strumenti per la qualità", definizione alquanto superata considerata l'evoluzione ed arricchimento di tali strumenti.

2. Il diagramma causa-effetto

Il diagramma causa effetto, rappresenta un efficace e pratico strumento per il problem solving.

Ogni output di processo è il risultato complesse relazioni di tipo causale fra più fattori. Nei processi non banali può risultare complicato l'esame simultaneo delle diverse cause che producono un certo effetto e delle relazioni fra queste.

Il diagramma causa-effetto è un semplice strumento grafico che permette di schematizzare i meccanismi di causa ed effetto che muovono un processo.

Per la sua forma è detto anche diagramma a lisca di pesce o diagramma di Ishikawa dal nome di chi lo utilizzò per la prima volta all'inizio degli anni '50.

Il J.I.S.¹ lo definisce come un diagramma che mostra le relazioni fra una caratteristica qualitativa ed i suoi fattori o cause.

Le caratteristiche che ne hanno favorito il vasto utilizzo nei settori più disparati sono:

- la *versatilità*: la logica organizzativa che ne è alla base lo rende un utile strumento per la risoluzione di problemi, anche molto complessi, in qualsiasi area produttiva, amministrativa o altro;
- la *semplicità*: riassume graficamente, in modo quindi diretto e comprensibile a tutti, relazioni anche complesse fra problemi e cause;
- permette una *visione d'insieme* del processo;
- favorisce l'*integrazione* delle attività di aree o settori diversi, semplificando i flussi comunicativi;
- evidenzia, fra le diverse cause, quelle che sono le più importanti e che necessitano di interventi prioritari;
- costituisce un punto di riferimento per attività di cooperazione;
- aiuta i neoassunti a comprendere le problematiche di un processo per loro nuovo, favorendone l'integrazione.

Per la sua compilazione sono possibili due approcci:

1) Prima procedura: *dal macrolivello al microlivello*.

Inizia con la determinazione del problema o della caratteristica qualitativa da analizzare (l'effetto).

Si stabilisce il team addetto alla compilazione. Dovrebbe essere il risultato del confronto fra tutti coloro (ingegneri, operai, responsabili di produzione...) che partecipano ad un certo processo (*brain storming*). Vengono tenute in

¹ Japanese Industrial Standards.

considerazione anche proposte fatte da persone non di specifica competenza di un particolare argomento o settore.

Si prendono delle pause di riflessione.

Si scrive la caratteristica qualitativa ad un'estremità di un foglio e la si circonda con un riquadro. Quindi si traccia una linea che parte dalla caratteristica e prosegue verso l'altra estremità.

Si ricercano quelle che sono le cause principali dell'effetto in esame e le si collegano, dopo averle riportate sul diagramma, all'effetto in esame tramite le cosiddette grandi lische.

Generalmente le cause principali di un processo appartengono a cinque categorie:

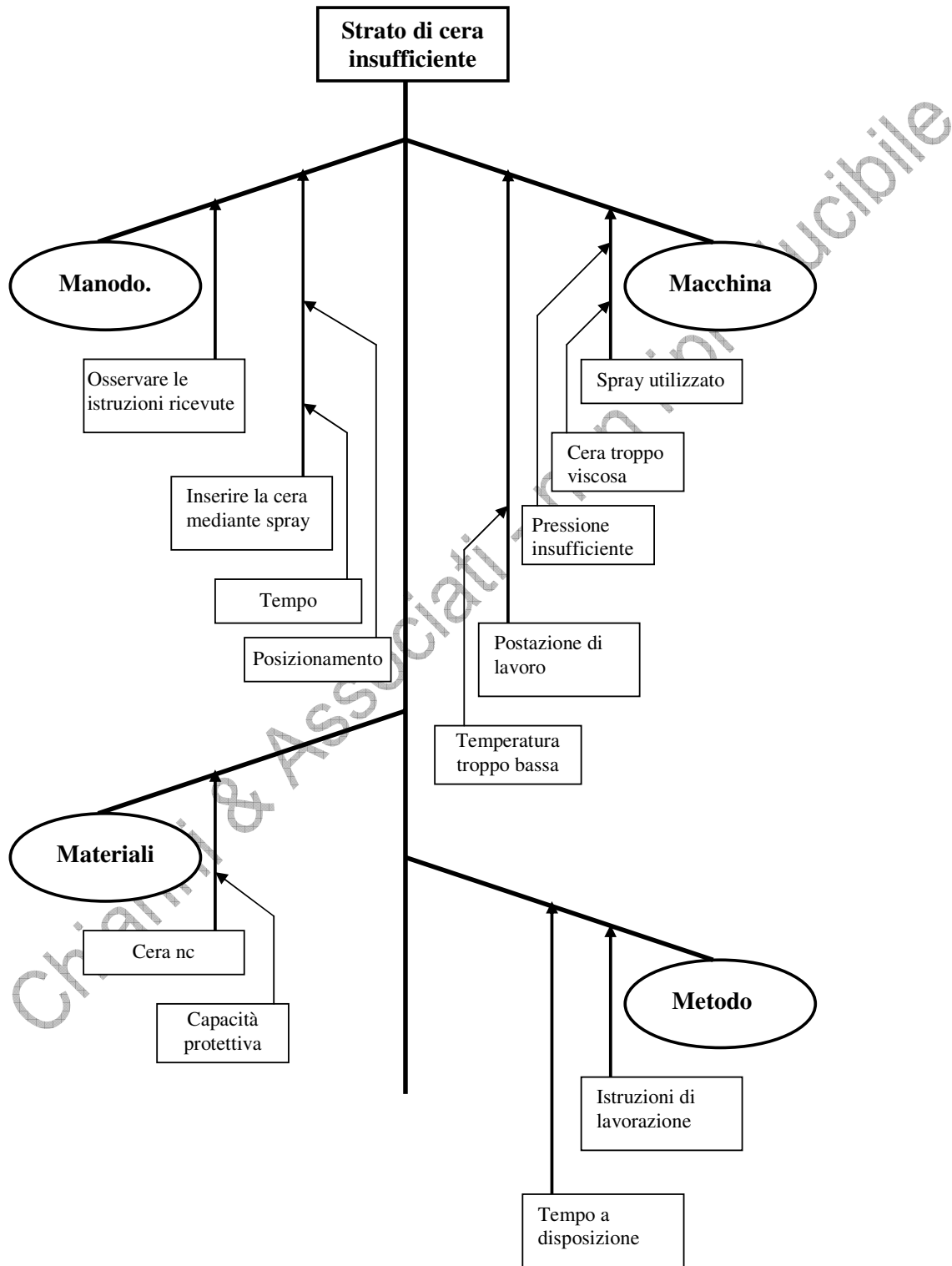
- materiali;
- macchinari;
- manodopera (personale);
- metodi;
- ambiente;

ma non si tratta di uno schema esaustivo. Lo schema classico di Ishikawa prevedeva quattro categorie, citate come *4M*:

- materiali;
- macchinari;
- manodopera (personale);
- metodi.

Non sono rare applicazioni con categorie ideate specificamente per il problema su cui indagare.

Esempio di diagramma causa-effetto (processo di applicazione manuale della cera all'interno del telaio per prevenire l'ossidazione)



Quindi si individuano le cause secondarie che influenzano le cause principali; anch'esse vengono riportate sul diagramma unendole a queste ultime con le cosiddette lische medie.

Analogamente si procede con le cause terziarie che influenzano le cause secondarie, collegate tramite le lische piccole.

Si può proseguire in questo modo per più livelli di analisi. È anche possibile indicare l'importanza, mediante un punteggio, delle diverse cause, al fine di evidenziare le priorità delle azioni correttive.

L'efficacia del diagramma di causa-effetto è direttamente proporzionale alla profondità dell'analisi. È quindi consigliabile dettagliare il processo il più possibile, esaminando anche i livelli infimi.

Dunque si analizza l'importanza relativa delle diverse cause, studiando il modo in cui queste influenzano l'effetto.

Le cause principali debbono essere ordinate in base a criteri non soggettivi, possibilmente quantitativi.

Si evidenziano le cause prioritarie sulle quali è necessario intervenire con sollecitazione.

A questo punto si registrano sul diagramma le altre informazioni necessarie come il titolo, il nome del prodotto o processo, le persone che vi partecipano, la data, etc.

Una volta completato il diagramma di Ishikawa è necessario riesaminarlo, correggerlo, approfondire il livello dell'analisi. Si deve fare un'analisi critica collettiva per verificare che tutte le possibili cause della variabilità siano state tenute in considerazione. Si studia ciò che è stato fatto e cosa sarà necessario fare per tenere sotto controllo ogni fattore.

Si deve cercare di migliorare il diagramma per risolvere con maggiore facilità le problematiche che lo hanno originato.

Ci si pone, in genere, una serie di domande: ci si chiede se esistano dei dati relativi al comportamento di una certa causa (ad esempio una Carta di Controllo che ne descriva l'andamento), se tale causa sia affetta da errori di misura o sistematici, se essa interagisca (ed in che modo) con altri fattori del processo.

2) seconda procedura: *dal microlivello al macrolivello*.

Per molti aspetti (finalità, approccio tramite brain storming, analisi finali) è simile al precedente. Tuttavia dopo l'individuazione della caratteristica qualitativa, ci si chiede quali fattori influenzino la variabilità di tale caratteristica e si elencano tutte le possibili cause per poi raggrupparle, dopo averne analizzato le interazioni, in cause di livello superiore, sino a giungere alle cause principali del processo.

Inversamente all'approccio precedente si parte dall'analisi delle piccole lische.

Suggerimenti per la costruzione dei diagrammi causa-effetto²:

- a) identificare tutti i fattori attraverso l'esame e la discussione fatta dalle persone coinvolte nel processo produttivo;
- b) esprimere le caratteristiche qualitative nel modo più concreto possibile;
- c) costruire tanti diagrammi causa-effetto quante sono le caratteristiche da studiare;
- d) scegliere, quando possibile, caratteristiche e fattori misurabili;
- e) individuare i fattori che consentono le azioni correttive;
- f) il grafico deve riflettere il più possibile il modo di pensare delle persone coinvolte nella sua costruzione.

² Fonte: Kume.

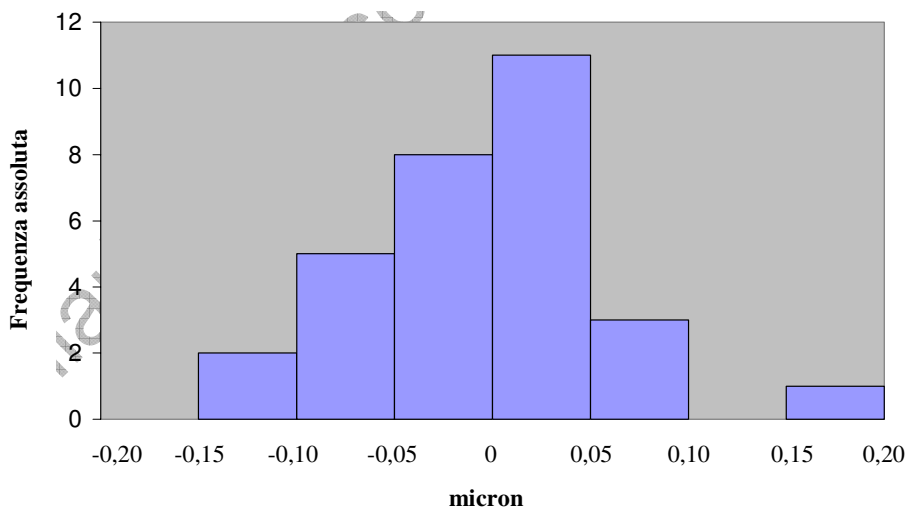
3. Istogramma di frequenza

Gli istogrammi (figura 2) sono un'altra tecnica grafica per la rappresentazione sintetica di dati; forniscono una rappresentazione dei parametri caratteristici di una distribuzione (costanti di posizione e di dispersione) ed inoltre ne forniscono anche la forma, utilizzata per capire con che tipo di distribuzione si ha a che fare.

Scarti dalla media di un processo in esame

Osservazione	Misura	Osservazione	Misura	Osservazione	Misura
1	0,02	11	-0,07	21	-0,14
2	-0,01	12	0,04	22	0,01
3	-0,11	13	0,01	23	0,06
4	-0,03	14	0,05	24	-0,04
5	0,02	15	-0,06	25	-0,02
6	0,08	16	0,01	26	-0,02
7	-0,02	17	0,04	27	0,01
8	0,17	18	-0,06	28	-0,07
9	-0,04	19	-0,03	29	-0,09
10	0,03	20	0,00	30	0,04

Istogramma relativo ai dati della tabella



Possono essere definiti come rappresentazione grafica di dati in ordine di grandezza in cui le frequenze osservate sono rappresentate da rettangoli la cui altezza corrisponde alla frequenza di ogni cella, intendendosi con quest'ultimo termine l'intervallo nella scala delle misure che corrisponde a ciascuna classe ordinata.

I valori possibili della distribuzione sono riportati su un asse (in genere l'ascissa) e le frequenze assolute (numero di osservazioni per un certo valore) o relative (rapporto fra frequenze assolute e numero di osservazioni) sull'altro.

Gli istogrammi sono un semplice modo per evidenziare le informazioni più importanti e lasciare da parte i dettagli inutili.

Per la costruzione di un istogramma:

- 1) si individuano il valore massimo X_{MAX} e minimo X_{min} di una distribuzione ed il range $X_{MAX}-X_{min}$;
- 2) si determinano il numero e l'ampiezza degli intervalli delle classi in cui vengono raggruppati i dati in modo tale che:
 - si abbiano da 4 a 20 celle;
 - gli intervalli siano di uguale ampiezza;
 - i limiti fra una classe e l'altra siano chiari e ben definiti (nessun dato può essere compreso in due classi);
 - si consiglia di usare un numero di intervalli che approssima la radice quadrata del numero n delle osservazioni (per $n >$ di 50) ;
- 3) si contano i dati presenti in ciascuna classe e si dispongono sul grafico dei rettangoli aventi per base le celle ed altezza proporzionale alla frequenza della classe.

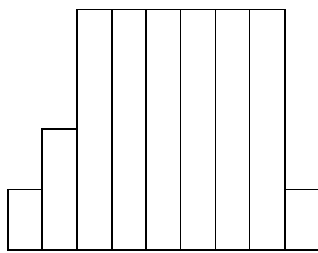
Si possono avere così utili informazioni relative alla distribuzione dei dati (unimodale, bimodale, asimmetrica, etc.).

Il confronto fra più istogrammi è molto semplice ed immediato.

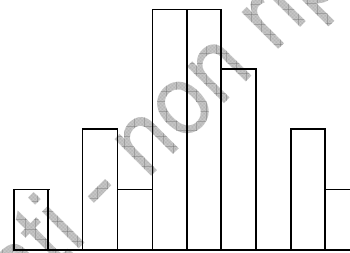
Se esistono dei limiti di specifica, devono essere riportati sull'istogramma per verificare se il processo sia capace (se nessun'unità prodotta cade al di fuori della regione delimitata dai limiti di specifica) o meno.

Prima di procedere nell'interpretazione dell'istogramma occorre, comunque, valutare la possibilità di errori nella raccolta dati. Si considerino, ad esempio, gli istogrammi sotto riportati nella figura 3.

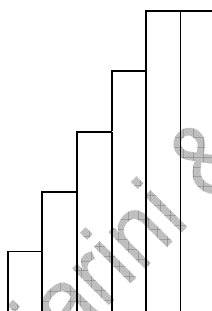
Istogrammi anomali



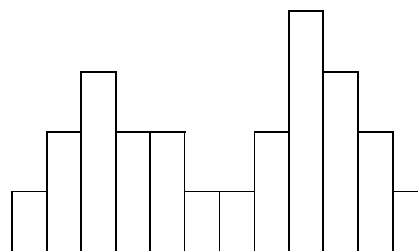
A - Rettangolare - piatto



B - A pettine



C - Assimetrico a precipizio



D - Bimodale

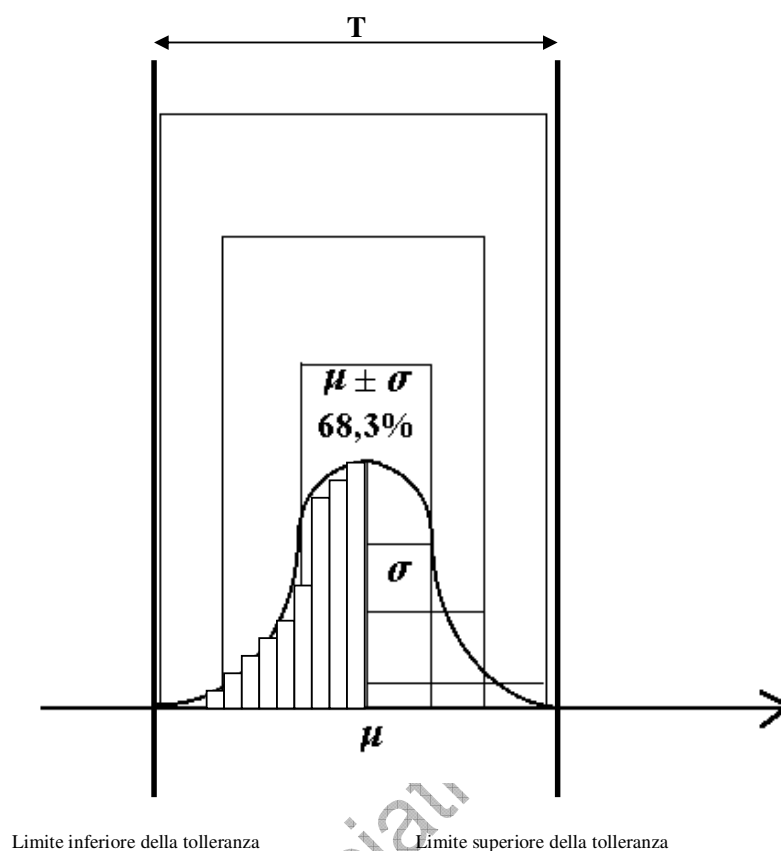
- Nel caso A si ha una distribuzione di tipo "piatto": è questo un classico errore dovuto all'uso di uno strumento di misura non idoneo (strumento non tarato, oppure a bassa precisione).

- Nel caso B si ha una distribuzione "a pettine": tipicamente questa distribuzione anomala la si ottiene quando il numero degli intervalli è troppo alto rispetto al numero dei dati.
- Nel caso C la distribuzione è asimmetrica, "a precipizio": ciò è normalmente legato ad errori nella raccolta dati (ad esempio un operatore che interpreta male le istruzioni e non registra i dati al di sopra di un certo valore). A volte, quando la asimmetria non è così spiccata come nel caso di figura C, può essere presente una sovrapposizione di dati non omogenei (processi, lotti, impianti, operatori, materiali, condizioni ambientali diverse, etc).
- Nel caso D la distribuzione è bimodale: l'anomalia è legata alla stratificazione di due campioni rappresentativi di situazioni diverse (processi, lotti, impianti, operatori, materiali, condizioni ambientali diverse, etc).

Valutata la presenza di distribuzioni anomale, come sopra, gli istogrammi possono essere utili per studiare preliminarmente i processi. Raccolto, infatti, un campione di dati da una popolazione di tipo normale, attraverso un'inferenza statistica, si possono fare delle previsioni sull'andamento del processo.

Come già visto nel capitolo 7 dedicato al Six Sigma e come si vedrà nel capitolo 9, attraverso indicatori di *capability* (capacità) si può mettere a confronto il range stabilito per il target (spesso citato come tolleranza) e la dispersione attorno al target.

Rapporto tolleranza e livello di sigma



Statisticamente, per misurare l'ottenimento del target, si usa un rapporto del tipo:

$$T/n\sigma,$$

con T = Tolleranza e σ = Deviazione Standard, che misura per l'appunto tale dispersione. Supponendo tale rapporto uguale a uno, possiamo dire che un "certo livello di sigma" (n volte σ) è contenuto nel range attorno al target. Dall'inferenza statistica è noto che, in queste condizioni, chiamata $F(z)$ la funzione cumulativa della distribuzione normale, la probabilità che un output del processo cada al di fuori della tolleranza è:

$$P = 1 - [F(n\sigma) - F(-n\sigma)].$$

Più n è alto e meno si ha probabilità che un output cada al di fuori della tolleranza, quindi che un prodotto - servizio risulti non conforme. In realtà, l'inferenza statistica, utilizza un campione, fotografia di breve periodo, per fare previsioni in un periodo più lungo nel quale, intervengono cause di variabilità nel processo. Per i processi che hanno a che fare con caratteristiche critiche o speciali, prima di avviare la produzione, in una fase cosiddetta di *validazione del processo produttivo*, si può studiare preliminarmente la *capacità della macchina*. L'obiettivo di questo studio è quello di indagare sulla sola variabilità della macchina, pertanto le restanti cause di variazione (si ricordino le 4 M) materiali, manodopera e metodi, devono essere sotto controllo. I passi di tale studio sono:

- 1) Si prende un campione di prodotti - servizi, prodotti con materiali conformi collaudati, un operatore addestrato e metodi corretti. L'obiettivo dello studio si ricorda essere la sola variabilità della macchina.
- 2) Si raccolgono minimo 30 valori come misure consecutive (osservazioni) sui prodotti - servizi della caratteristica critica (es. un diametro, una temperatura, un tempo, etc.).
- 3) Si costruisce l'istogramma tramite le indicazioni sopra.
- 4) Si valutano le anomalie di distribuzione come riportato in figura 3.
- 5) Si valuta la normalità del campione sia in modo visivo, sia tramite indici di posizione quali la *media*, *moda* e *mediana*.

- La *media* è data dalla somma dei valori osservati divisa la numerosità del campione n ;

- la *moda* è il valore della caratteristica che ha la maggior frequenza (ordinata più alta);

- la *mediana* è quel valore della caratteristica per cui metà dei valori sono più grandi e metà più piccoli della mediana.

In una distribuzione normale *media*, *moda* e *mediana* coincidono.

6) Si calcola la deviazione standard
$$\sigma = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2\right)}$$
,

dove n indica la numerosità del campione, x_i l' i -esimo valore, e $\mu = \bar{X}_i$ la media dei valori del campione;

7) Si calcolano gli indici:

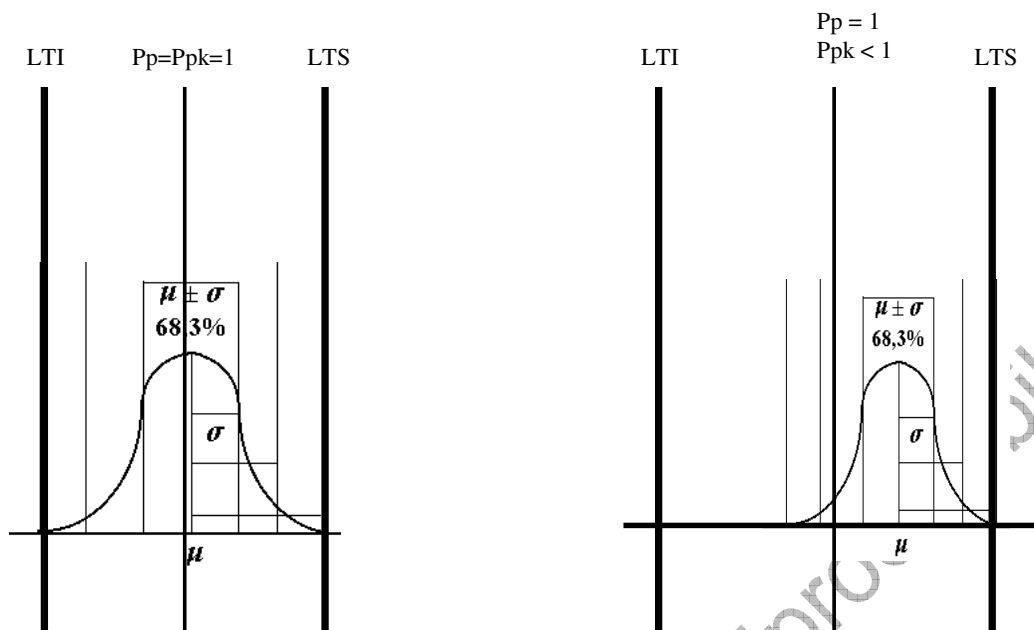
$$Pp = \frac{LTS - LTI}{6\sigma}$$

$$Ppk = \min\left\{\frac{LTS - \mu}{3\sigma}, \frac{\mu - LTI}{3\sigma}\right\}$$

dove LTS è il limite superiore della tolleranza ed LTI il limite inferiore.

Pp è la misura del rapporto tra la dispersione ammissibile per la macchina, calcolata dalla differenza tra i Limiti di tolleranza e la dispersione reale, rappresentata dal valore 6σ , detta anche *Tolleranza Naturale*, TN . Si utilizza 6σ perché in una distribuzione normale, come si suppone che sia quella sotto esame, il 99,73% degli elementi è compreso in sei volte la deviazione standard.

Ppk misura, invece, sia la dispersione che la centratura della macchina, tenendo conto dell'ampiezza della distribuzione e della posizione in cui è posta, rispetto al punto medio della tolleranza. Scegliendo il minore dei due valori calcolati, si determina quanto è capace la macchina sul lato peggiore, quello cioè rappresentato dalla coda della gaussiana più vicina al limite di tolleranza.



Se P_p è maggiore di 1, l'ampiezza 6σ dei dati cade completamente nei Limiti di Tolleranza; un P_p compreso tra 0 e 1 indica che una parte dei prodotti del processo cadono oltre i Limiti, ed infine un P_p negativo indica che la media dei dati non è nella tolleranza. Un P_p uguale ad 1 indica che il 99,73% delle parti prodotte è nei limiti di tolleranza, cioè devono essere respinte solo 3 parti su 1000. Anche se apparentemente un valore di P_{pk} di 1 sembra il minimo accettabile, nella pratica si consiglia tuttavia di mantenere tale valore a 1,33 (che equivale ad ottenere $T/8\sigma = 1$), o addirittura a 1,67, in modo da premunirsi contro l'insorgenza di eventuali cause speciali successivamente in produzione.

Se il risultato dello studio non è positivo, si agisce sulla macchina e si riparte dal punto 1 con un nuovo campione.

- 8) Validata la variabilità della macchina ed avviata la produzione, occorre tenere sotto controllo il processo. Lo strumento tipico è la *carta di controllo*. La deviazione standard calcolata nello studio preliminare di capacità di macchina, può essere utilizzata per calcolare i limiti di controllo superiore ed inferiore della primissima carta.

Si fa notare che i passo sopra sono validi per un campione iniziale nel quale si sono utilizzati materiali conformi, operatori addestrati e metodi certi, studiando la variabilità di processo legata alla sola macchina. Nella normale e successiva produzione, attraverso le carte di controllo, si valutano tutte le cause di variabilità sul processo.

3.1 Grafico rami e foglie

Il *grafico rami e foglie* (stem and leaf plot), proposto da Taguchi nel 1977, costituisce un metodo alternativo per rappresentare un istogramma con dati raggruppati in classi. È una sintesi fra tabella e grafico: ha i valori numerici espressi tramite cifre ed il suo profilo è simile a quello di un istogramma.

Si suddividono le cifre dei dati in due parti: una formata dalla o dalle cifre iniziali (quelle a sinistra) e l'altra dalle restanti (a destra). Le prime cifre costituiscono i rami, quelle più a destra le foglie.

Si stabilisce il numero dei rami (in genere fra i 5 e i 20), quindi si collocano sul margine sinistro del grafico le cifre corrispondenti a ciascuno di essi. A fianco di ciascun ramo si collocano le foglie corrispondenti nell'ordine in cui i dati sono stati rilevati, oppure ordinate in senso crescente.

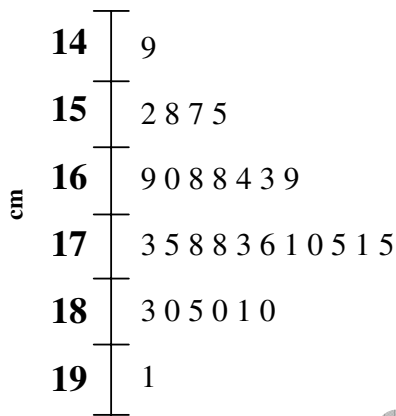
Il grafico rami e foglie ha i seguenti vantaggi:

- non comporta la scelta del numero ed ampiezza delle classi, ma solo dei rami;
- conserva i valori numerici delle singole osservazioni;
- se i dati nelle foglie sono ordinati in senso crescente, permette una semplice identificazione dei quantili.

Tabella 2: Altezza dei dipendenti

Dipendente	Altezza (Cm)	Dipendente	Altezza (Cm)	Dipendente	Altezza (Cm)
1	183	11	191	21	164
2	173	12	180	22	175
3	152	13	168	23	157
4	175	14	173	24	172
5	169	15	176	25	180
6	149	16	181	26	155
7	180	17	168	27	163
8	185	18	171	28	171
9	178	19	170	29	175
10	160	20	158	30	169

Figura 4: Esempio di grafico rami e foglie normale per la tabella 2



Nel grafico *rami e foglie stirato*, variante del normale rami e foglie, ogni ramo viene suddiviso in due parti: una comprende i numeri (foglie) da 0 a 4, la seconda da 5 a 9. Similmente il grafico *rami e foglie compresso* ha cinque rami al posto di uno del normale rami e foglie: il primo indicato con * o ' comprende le cifre 0 e 1, il secondo, indicato con t o ", le cifre 2 e 3, il terzo, f o ""', contiene il 4 ed il 5, il quarto, s o """, 6 e 7, l'ultimo, • o """, 8 e 9.

Per la scelta del grafico più appropriato fra le tre varianti (normale, compresso e stirato) si calcola:

$$L=10\log_{10}n$$

dove n rappresenta il numero delle osservazioni ed L il numero massimo di rami del grafico.

4. Il diagramma di Pareto

Le problematiche della qualità costituiscono un costo per le aziende. È importante analizzare la distribuzione di questi costi in relazione alle loro cause. In genere la maggior parte delle perdite economiche è dovuta a poche cause attribuibili (definite “the vital few”), mentre la maggior parte delle cause produce un numero limitato di effetti. Questa legge, nota come *principio di Pareto*³, aiuta a focalizzare l’attenzione e le risorse aziendali su quelle poche cause attribuibili che sono responsabili della maggior parte dei costi (le variabili critiche di un processo), trascurando, almeno in un primo momento, le altre cause meno importanti.

Il principio di Pareto deriva dagli studi effettuati alla fine del secolo scorso dall’economista che gli ha dato il nome sulla distribuzione del reddito. Pareto notò che meno del 20% della popolazione possedeva oltre l’80% delle risorse.

Successivamente Lorenz – economista americano – rappresentò questo principio tramite un diagramma (il diagramma di Pareto).

Juran riprese il principio nell’ambito del controllo della variabilità di un processo.

Il diagramma di Pareto è uno strumento di sintesi grafica che permette di evidenziare “the vital few”, le poche variabili critiche di un processo che necessitano di priorità assoluta.

Le fasi per la sua costruzione sono:

- 1) si stabilisce il problema da esaminare e se ne individuano le cause (cercando di non tralasciarne alcuna);
- 2) si verifica, in un lasso di tempo prestabilito, con quale frequenza ciascuna causa si verifica;
- 3) per ciascuna causa si calcola in un’apposita tabella la frequenza assoluta e relativa;

³ Dal nome del celebre economista italiano (1848-1923).

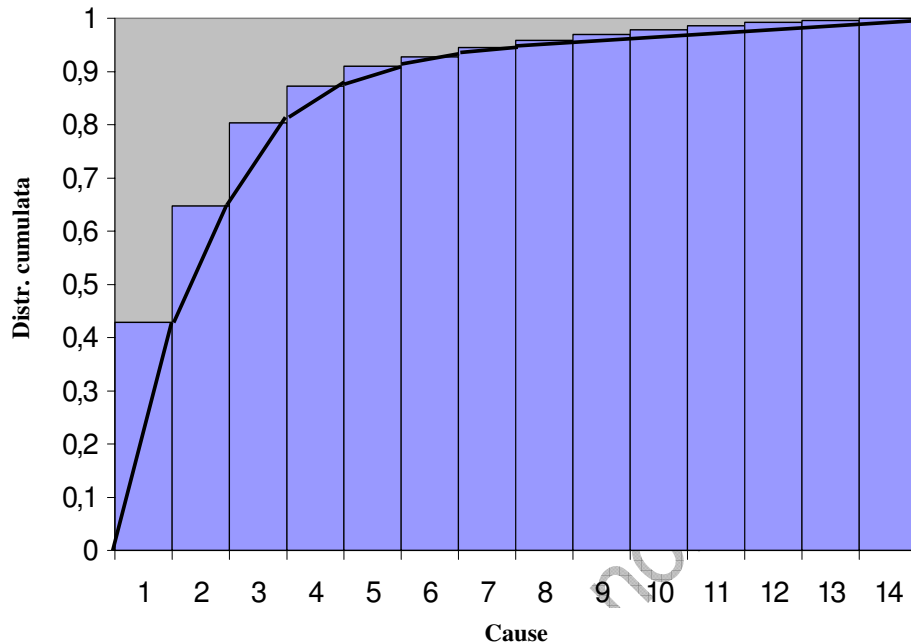
- 4) si ordinano le cause contenute nella tabella in senso decrescente. La prima causa sarà quella con frequenza maggiore;
- 5) si calcola la distribuzione cumulata delle cause in termini assoluti e percentuali (tabella 3);

Componenti difettose su 800 computer esaminati.

N°	Componente	Freq. ass.	Freq. rel.	Distr.cum.
1	Lettore CD/DVD	343	0,42875	0,42875
2	Tastiera	175	0,21875	0,64750
3	Scheda video	125	0,15625	0,80375
4	ULA	55	0,06875	0,87250
5	Trasformatore	30	0,03750	0,91000
6	Cache	14	0,01750	0,92750
7	Porte	14	0,01750	0,94500
8	Modem	11	0,01375	0,95875
9	Pulsanti	9	0,01125	0,97000
10	Processore	7	0,00875	0,97875
11	Hard disk	6	0,00750	0,98625
12	Porta USB	5	0,00625	0,99250
13	Scheda madre	3	0,00375	0,99625
14	RAM	3	0,00375	1

- 6) si crea lo schema per il diagramma di Pareto: è costituito da un quadrato, sul lato sinistro del quale si riporta una scala che va da 0 al totale della distribuzione e sul lato destro la scala percentuale della distribuzione cumulata. Sull'asse orizzontale si riportano le diverse cause in ordine decrescente per frequenza;
- 7) all'interno del quadrato si costruisce un istogramma che rappresenti la distribuzione delle frequenze delle diverse cause;
- 8) a questo punto si riporta all'interno del quadrato la cosiddetta curva cumulata di Pareto, la quale rappresenta, in corrispondenza di ogni causa, la percentuale cumulata corrispondente. I punti così ottenuti vengono poi uniti con una linea;

Diagramma di Pareto per i dati in tabella



9) si inseriscono le informazioni ausiliarie quali il titolo, il nome del processo analizzato, il periodo di tempo, il numero totale delle informazioni, le persone che vi hanno lavorato, etc.

10) si analizza il diagramma. Se non si sono evidenziate “the vital few” (la curva di Pareto è molto vicina alla diagonale del diagramma) è consigliabile provare a classificare le cause in modo diverso.

Esistono due tipi di diagramma di Pareto:

- a) diagramma di Pareto per cause: serve per individuare le principali cause della difettosità che, in genere, vengono catalogate per operatore, macchina, materie prime e metodi;
- b) diagramma di Pareto per fenomeni: finalizzato ad individuare la difettosità in termini di qualità, costi, consegna e sicurezza.

5. Il foglio di raccolta dei dati

Alla base di ogni procedura di misurazione ed analisi c'è un insieme di dati. Queste informazioni devono possedere le seguenti caratteristiche:

- la validità dei dati deve essere elevata (almeno del 95%);
- deve essere possibile ricostruire l'andamento temporale dei dati raccolti;
- il tipo di dati raccolti deve essere appropriato, adatto alle esigenze. Per questo si utilizzano strumenti quali il diagramma di causa-effetto e di Pareto;
- i dati devono essere rilevati in tempo reale ed in linea.

Il foglio di raccolta dati è uno strumento che serve per semplificare e rendere più affidabile il processo di raccolta delle informazioni.

Non è uno strumento standard, ma deve essere progettato ad hoc. Deve dare la possibilità aggregare i dati automaticamente in modo da semplificare e velocizzare le elaborazioni successive.

Spesso ad un processo di raccolta dati collaborano e/o si alternano diverse persone. Il foglio deve uniformare l'operato dei diversi individui coinvolti.

Il seguente foglio (Figura 6) può essere utilizzato per controllare la distribuzione del parametro di un certo prodotto, ad esempio il diametro di un bullone con valore medio 10,00.

Foglio di raccolta dati per le dimensioni del diametro del bullone metallico.

Periodo: Gennaio Operatore: Rossi Andrea.

<u>Misura</u>	<u>Registrazioni</u>	<u>Totale osservazioni</u>
10,05		
10,04	X	1
10,03		
10,02	XXX	3
10,01	XXXXXXXXXX	15
10,00	XXXXXXXXXX	26
9,99	XXXXXXXXXX	9
9,98	X	1
9,97		
9,96		
9,95		
		TOTALE: 55

6. Il diagramma di correlazione o grafico a dispersione.

Il diagramma di correlazione è utilizzato per analizzare le relazioni fra due variabili X e Y.

Per la sua costruzione:

- a) Si definiscono le due variabili delle quali si intende studiare la relazione e si raccolgono (almeno una trentina) le coppie di valori (X_i, Y_i) che generalmente vengono riportate in una tabella a due entrate come la seguente tabella 4;
- b) Tenendo in considerazione i valori minimi e massimi delle due distribuzioni si stabiliscono le unità di misura dei due assi di un piano cartesiano, in modo tale che la lunghezza di ascissa e ordinata sia simile. Se una delle due variabili dipende dall'altra, si riporta la variabile dipendente sull'ordinata e la variabile indipendente sull'ascissa.

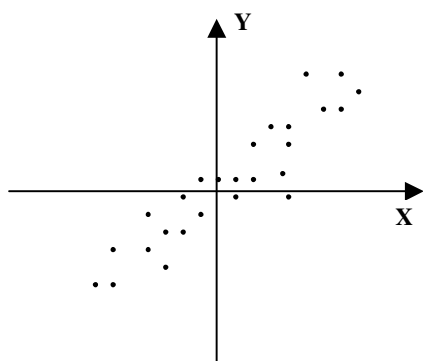
Tabella 4: Osservazioni delle variabili X ed Y

X_i	Y_i

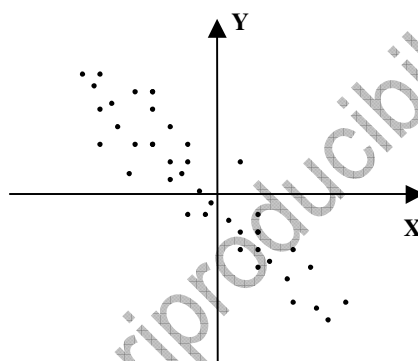
- c) Dopo aver disegnato gli assi con le rispettive unità di misura, si riportano sul piano cartesiano i punti corrispondenti alle coppie di valori (X_i, Y_i) . Se risultano dei valori identici per più coppie si utilizzano dei punti più marcati.

d) Si riportano sul grafico le informazioni supplementari quali il periodo di osservazione, il nome delle variabili osservate, le unità di misura, la numerosità campionaria, il titolo del grafico, il nome dei compilatori, etc.

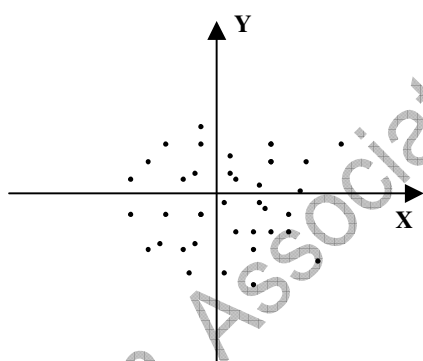
Diagrammi di correlazione



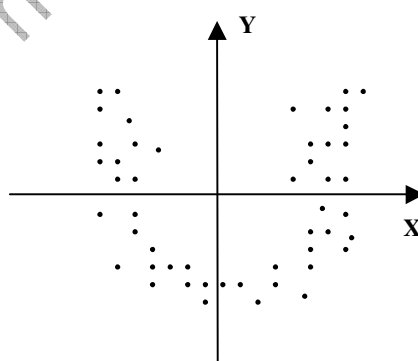
Correlazione lineare positiva



Correlazione lineare negativa



Assenza di correlazione lineare



Correlazione di secondo grado

Partendo dalla disposizione dei punti sul grafico, si analizza la relazione fra le variabili.

Se i punti appaiono disposti casualmente non vi è motivo di credere che esista una relazione fra le variabili, se invece i punti tendono a disporsi lungo una retta esiste correlazione lineare fra le variabili, se i punti tendono ad assumere la forma di una funzione di grado superiore al primo, come ad esempio una parabola, esiste una correlazione non lineare fra le variabili.

Si deve anche interpretare l'eventuale presenza di punti anomali, isolati dagli altri. Possono essere il frutto di un errore di misurazione ed in tal caso non

debbono essere tenuti in considerazione, oppure possono essere dovuti ad una qualche causa particolare che deve essere individuata. In quest'ultimo caso i punti non devono essere ignorati, ma si deve capire cosa li ha prodotti.

Quando fra i dati è presente un fattore di stratificazione è consigliabile raffigurare i dati stratificati con colori o simboli diversi, in base alla classe di stratificazione di appartenenza.

Il grafico a dispersione ha il vantaggio di evidenziare tutti i valori rilevati.

Come precedentemente detto, nel caso di correlazione lineare per quantificare la

$$-1 \leq r = \frac{\text{Cov}(X, Y)}{\sqrt{\text{VAR}(X)\text{VAR}(Y)}} \leq +1$$

relazione espressa graficamente si utilizza il coefficiente di correlazione lineare di Bravais Pearson, dato da:

dove se:

$r=+1$: perfetta dipendenza lineare diretta (retta con inclinazione positiva);

$r=0$: assenza di correlazione;

$r=-1$: perfetta dipendenza lineare inversa (retta con inclinazione negativa);

La correlazione non implica necessariamente un rapporto di causalità fra le due variabili: entrambe, per esempio, possono essere legate ad una terza, causa di entrambe.

Bisogna anche fare attenzione alla falsa correlazione: può capitare di trovare un valore di $|r|$ elevato fra due variabili, fra le quali non c'è in realtà alcun legame.

Oltre al coefficiente di correlazione lineare si è soliti calcolare e riportare sul diagramma le rette di regressione, per quantificare la relazione fra le variabili mediante una funzione matematica (se esiste).