Dicembre, 2009

L'analisi dei rischi: l'aspetto statistico

Ing. Pier Giorgio DELLA ROLE – Six Sigma Master Black Belt

Introduzione

Nell'esecuzione dei progetti Six Sigma è di fondamentale importanza sapere se i risultati ottenuti a seguito di un cambiamento nei processi sia attribuibile al caso oppure se la differenza sia realmente significativa e quindi attribuibile alle azioni intraprese.

Errore alfa, errore beta e potenza del test

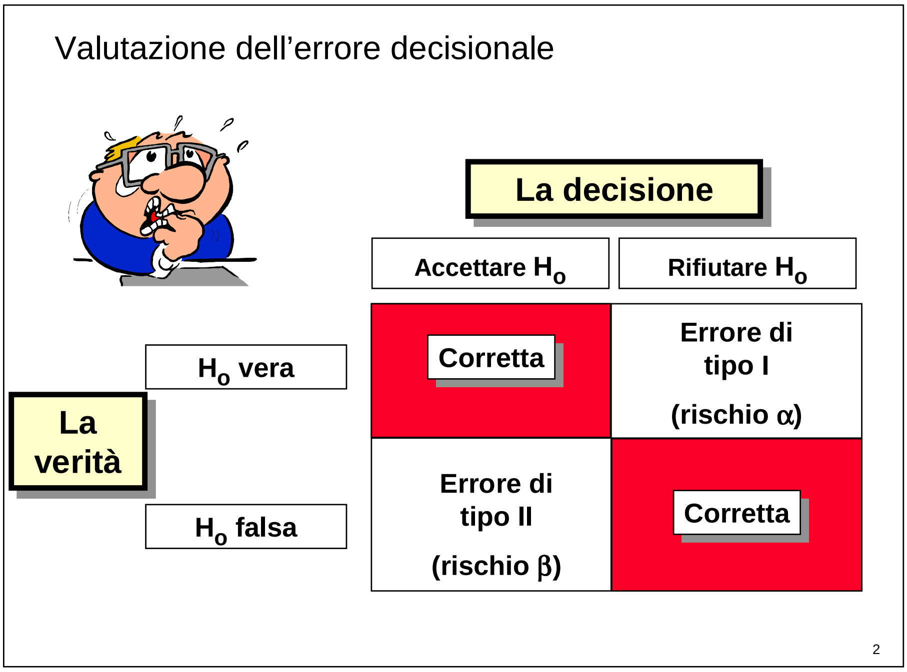
Si parte dal concetto di variabilità che si fonda sul fatto che non esistono due oggetti perfettamente uguali, per quanto elevata sia stata la cura con cui sono stati realizzati.

Il problema è sapere se tali differenze osservate siano attribuibili alla sola variabilità naturale (in pratica al caso), oppure la differenza sia realmente significativa e quindi dovuta a cause speciali (azioni intraprese per migliorare le performance di un processo).

Quando si deve giudicare la diversità di due gruppi di dati (in pratica i dati relativi ad un processo prima dei miglioramenti introdotti e i dati misurati dopo i miglioramenti) si corrono *due tipi di rischio* e quindi le combinazioni possibili sono quattro.

Una considerazione di carattere generale è la seguente: poiché i test statistici lavorano sulle medie dei gruppi o campioni da confrontare, sono preferibili campioni di numerosità significativa in quanto la dispersione delle medie campionarie diminuisce all'aumentare della numerosità del campione (teorema del Limite Centrale).

La tavola delle decisioni (truth table) e il Test delle Ipotesi



I **Test delle Ipotesi** mettono a confronto due ipotesi:

- **Ipotesi Nulla : Ho**
Ipotesi di non differenza tra i gruppi (presenza di sole cause naturali)

- **Ipotesi Alternativa : Ha**
Ipotesi di gruppi significativamente differenti (presenza di cause specifiche)

Vediamo come funzionano i Test delle Ipotesi.

L'ipotesi nulla Ho è per definizione sempre la più conservativa; per esempio in un test fra due processi (fra prima e dopo il miglioramento) l'ipotesi nulla prevede che non vi sia differenza cioè in pratica che l'effetto medio del processo migliorato sia pari a quello ante modifica.
I test statistici (tutti i t-test, Anova,…) devono tentare di confutare questa ipotesi, dimostrando al contrario un'ipotesi alternativa Ha (c'è differenza fra gli effetti medi dei due processi).
In realtà i valori misurati quasi mai coincideranno (concetto di variabilità espresso all'inizio) tra i due gruppi. E' infatti poco probabile la perfetta uguaglianza di 2 valori che, anche solo per puro effetto del caso, tenderanno invece ad essere leggermente diversi tra loro.
Come riconoscere allora gli effetti del caso da quelli veri, dovuti cioè ad una reale efficacia dell'intervento di miglioramento?

Grazie alle tecniche statistiche, è possibile associare ad ogni differenza osservata una ***probabilità*** che tale differenza si verifichi. Si stabilisce convenzionalmente una soglia di probabilità, solitamente il 5%, al di sotto della quale si prenderà la decisione di rifiutare l'ipotesi nulla.

Questa probabilità è detta ***alfa o livello di significatività*** e rappresenta dunque il rischio, che si accetta a priori di correre, di cadere nell'errore di primo tipo, di assumere cioè come frutto di un miglioramento una differenza che in realtà è frutto solo del caso.

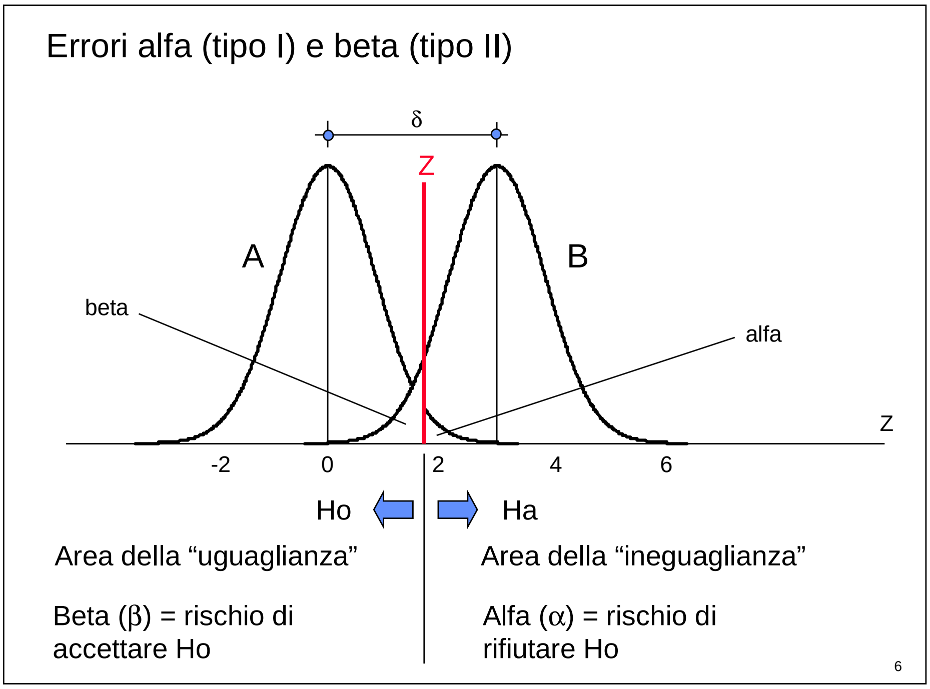
Il complemento a 1 di alfa, 1 – alfa, si chiama ***livello di fiducia.***
Per esempio se alfa = 0,05, 1 – alfa = 0,95 (95%).

Se si decidesse, per proteggersi dall'errore di primo tipo, di limitare al massimo il livello di alfa, si correrebbe il rischio di commettere l'errore opposto: accettare l'ipotesi nulla quando essa è falsa o, in parole più semplici, non vedere una differenza reale tra i due gruppi.

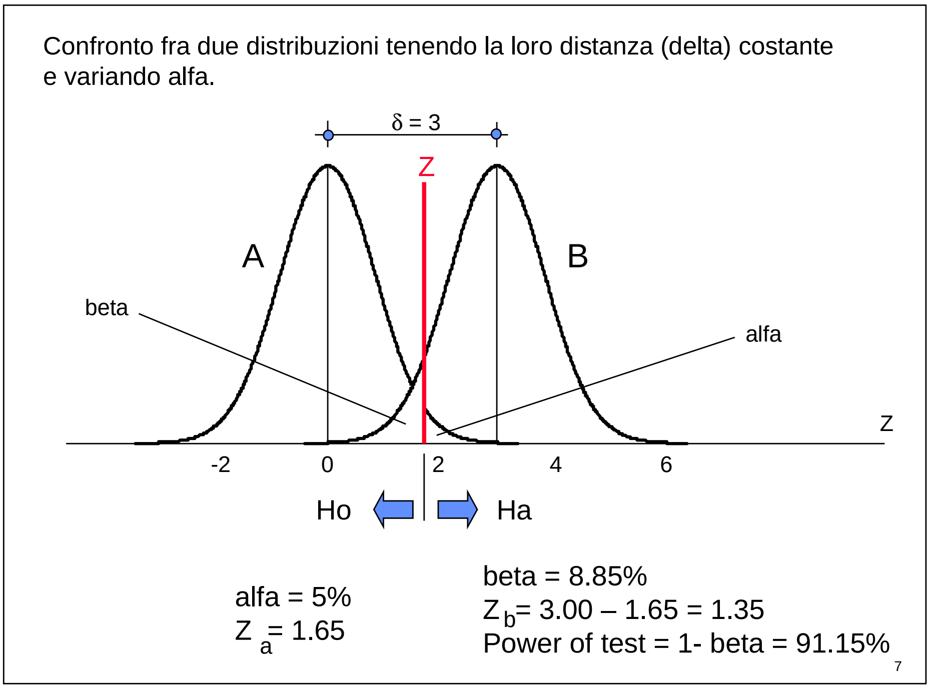
Anche in questo caso si verificherebbe un errore, detto di ***secondo tipo***.
La probabilità di commettere questo errore è detta ***beta***.

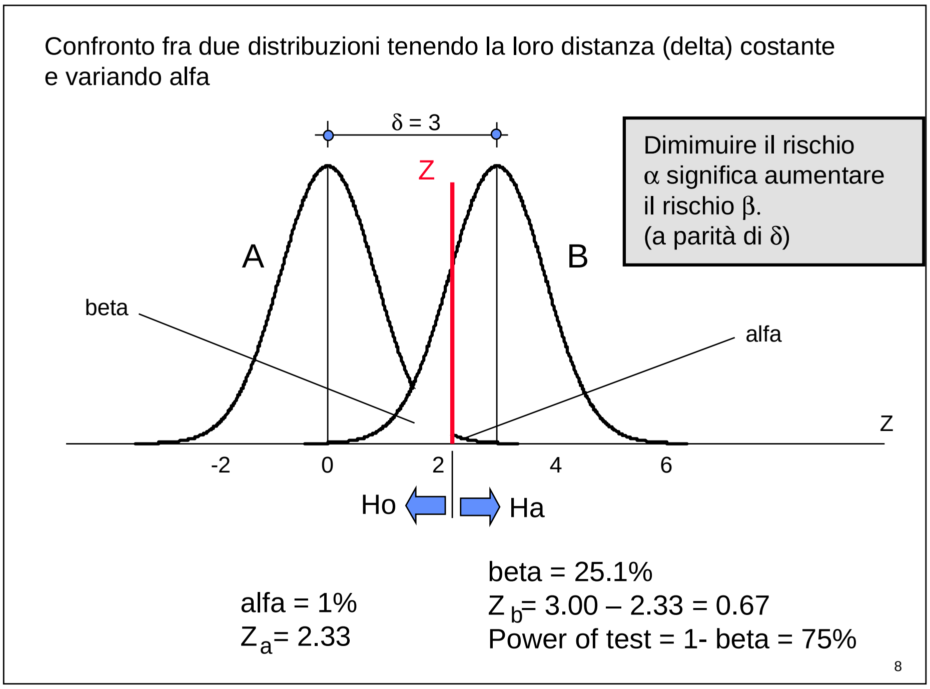
Il suo complemento a 1,1 – beta, esprime la ***potenza del test*** e, in effetti misura il potere che un test ha di svelare un effetto esistente.
Per esempio per beta = 0,2 , 1 –beta = 0,8 (80%).

La figura sottostante mostra graficamente gli errori alfa e beta supponendo gli output di due processi A e B distribuiti normalmente e con uguale varianza.



A parità di ***delta*** (differenza fra le medie dei due processi), diminuire il rischio alfa significa aumentare il rischio beta cioè la possibilità si scoprire la differenza delta esistente.





Nelle applicazioni Six Sigma per migliorare un processo, occorre quindi decidere in anticipo il livello di alfa e di beta e di conseguenza si predeterminerà la numerosità dei campioni da ***misurare per ogni processo*** (prima e dopo il miglioramento).

Dopo l'esecuzione del test, si misura sui dati ottenuti la probabilità, ***detta p-value***, di ottenere i dati osservati sotto ipotesi nulla: se il valore di p-value è inferiore al livello prefissato di alfa (per esempio $p < 0,05$) si rifiuta l'ipotesi nulla (attribuendo perciò le differenze osservate alle azioni di miglioramento) affermando che l'effetto è statisticamente significativo; altrimenti non la si rifiuta, attribuendo le differenze osservate al caso.

In entrambe le condizioni si è consapevoli della probabilità di errore della decisione, avendola fissata a priori.

Parlando di miglioramento di un processo secondo la terminologia Six Sigma
(Y = f(X)) si può concludere che:

. Errore alfa = significa agire su di una variabile X che non ha effetti sul risultato Y.
Valori tipici di alfa sono: 0,05 – 0,10

. Errore beta = non agire su di una variabile X significativa cioè trascurare una opportunità di miglioramento.
Valori tipici di beta sono: 0,10 – 0,20

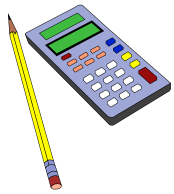
Dimensione del Campione (Sample Size)

Per poter calcolare correttamente la dimensione del campione dobbiamo definire le quattro variabili che compaiono nella formula (vedi figura):

Calcolare Sample Size

- La relazione tra le cinque variabili α, β , δ, σ e N è espressa dalla seguente equazione:

$$\boxed{N = \frac{2(Z_{\alpha/2} + Z_{\beta})^2}{\delta^2 / \sigma^2}}$$



- Se conosciamo quattro variabili, possiamo calcolare la quinta.
- Usando Minitab, possiamo calcolare ‘sample size‘ e ‘power of test‘:
  - Sample size = N.
  - Statistical power = 1–β.

12

Abbiamo parlato degli errori alfa e beta che entrano nella formula, vediamo adesso di dire qualcosa sulle altre due variabili : ***delta e sigma.***

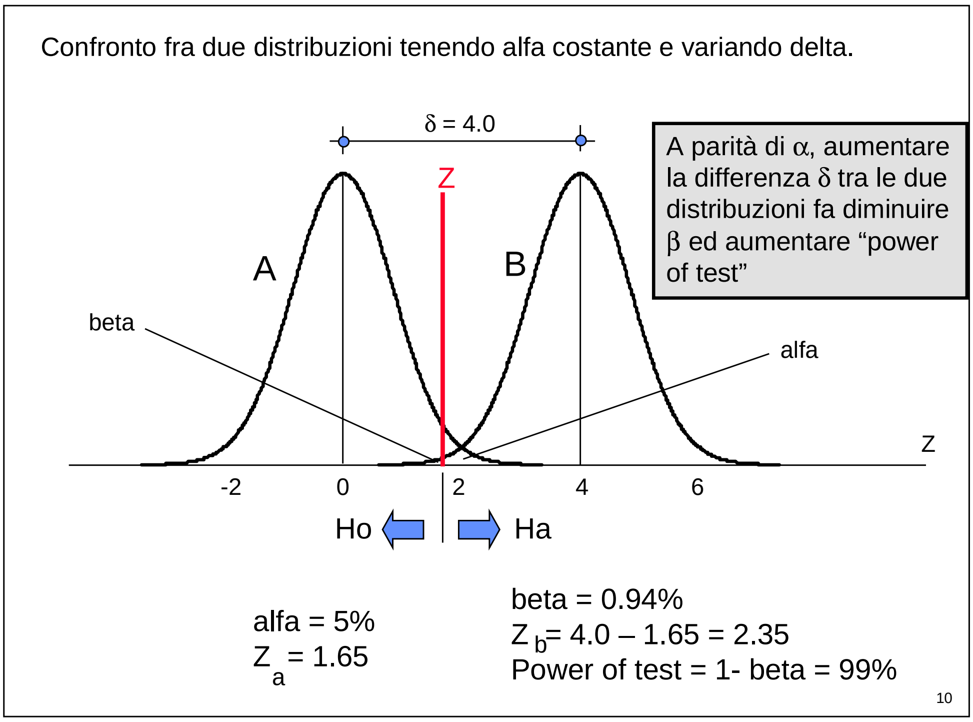
- Delta è la misura del cambiamento minimo (effetto) che vogliamo essere ragionevolmente sicuri di trovare durante il test; è nota anche come ***segnale***;

- Sigma è la deviazione standard del processo; è anche conosciuta come ***rumore***;

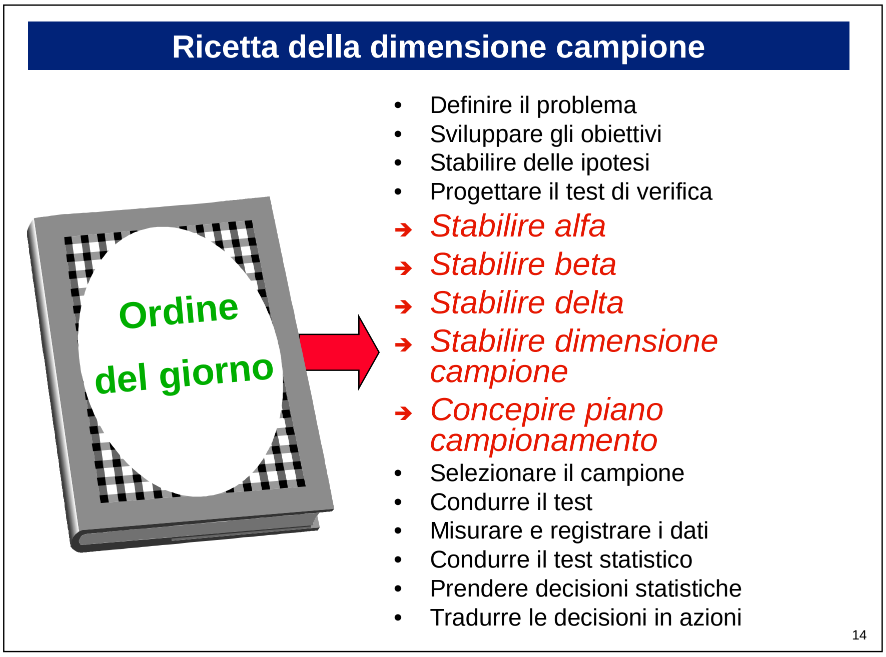
- Il rapporto “segnale/rumore” è direttamente collegato alla nostra capacità di scoprire un cambiamento nel processo.

Gli alti rapporti segnale/rumore sono associati ad una buona potenza del test (= 1 – beta), cioè alla capacità di scoprire un cambiamento quando questo si verifica.

Nella figura sottostante infatti si è variato “delta” a parità di “sigma” e quindi il rapporto segnale/rumore è aumentato.



La ricetta per una corretta "dimensione del campione" risulta pertanto essere:



La parte di calcolo può essere facilmente eseguita con ***Minitab***.

I comandi sono sotto *"Stat → Power and Sample Size"* poi, dal momento che stiamo a tutti gli effetti utilizzando un Test delle Ipotesi, si tratta di scegliere il tipo di test più appropriato.

Facciamo un esempio

Un team ha sviluppato una nuova procedura per la gestione dei test in un laboratorio della Qualità.
Il team è pronto a fare un pilota della nuova procedura e vuole sapere quanti campioni sono necessari per scoprire se c'è differenza tra le due procedure (vecchia e nuova).

Si tratta quindi di un 2 sample t-test.

Dati in Minitab:

- Standard deviation (sigma) : 10 minuti
- Size of difference (delta) : 12 minuti
- Significance level : 0,05
- Power of test (1 – beta) : 0,90 (90%)

Risultati da Minitab:

Risultati (from Minitab)

Power and Sample Size

```
2-Sample t Test

Testing mean 1 = mean 2 (versus not =)
Calculating power for mean 1 = mean 2 + difference
Alpha = 0,05  Assumed standard deviation = 10

            Sample  Target
Difference    Size   Power  Actual Power
        12      16     0,9      0,907190
```

Probabilità di scoprire la differenza di 12, se esiste, usando 16 campioni in ogni gruppo

Risposta: 16 campioni sono richiesti per i 2 gruppi

501

La risposta è che sono necessari 16 misure relative alla procedura prima del miglioramento e 16 misure sulla procedura pilota per scoprire, se esiste, una differenza di 12 minuti.
Minitab calcola poi il valore effettivo della potenza del test pari a 0.907.