



Analisi della Varianza ANOVA

Nasce negli anni '30 per problemi relativi l'agricoltura

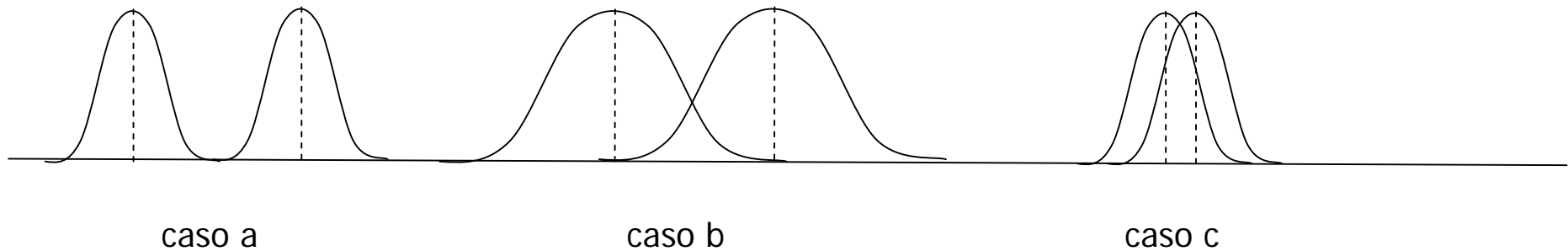
Quando si indaga un processo è impossibile eliminare la variabilità del caso

Mediante l'ANOVA è possibile scomporre la variabilità totale
nelle componenti imputabili ai fattori e nelle componenti imputabili al caso



Assunzioni

Ognuno dei gruppi di osservazioni deve essere un campione casuale che proviene da una popolazione di tipo normale. I campioni devono essere indipendenti tra di loro



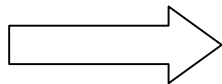
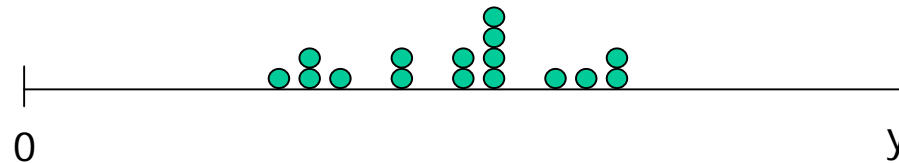


Fattore: origine della variazione

Livello: valore assunto dal fattore al momento della misura

y_i : valore della misura

Può essere considerata come un punto materiale di peso y_i a distanza y_i dall'origine



$$\sum y_i^2$$

Momento rispetto all'origine



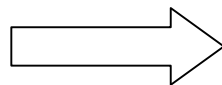
Il momento rispetto all'origine del sistema di punti è uguale al momento rispetto all'origine del peso complessivo concentrato nel baricentro (valore medio) aumentato della somma dei momenti di ciascun punto materiale rispetto al baricentro

$$\sum y_i^2 = \sum (y_i - \bar{y})^2 + n\bar{y}^2$$

Somma di tutti i dati

Correzione per la media

$$n\bar{y}^2 = n\bar{y} \cdot \bar{y} = n \frac{\sum y}{n} \frac{\sum y}{n} = \frac{T^2}{n} = n\bar{y} \frac{\sum y}{n} = \bar{y} \sum y = CM$$



$$\sum (y_i - \bar{y})^2 = \sum y_i^2 - CM$$

Divisa per i GdL fornisce la varianza (MS)

Somma dei quadrati (QT o SST)



Esempio 1: un solo fattore controllato

% utilizzo CPU	Temperatura CPU
0	19
0	17
0	16
20	20
20	23
20	21
50	25
50	22
50	26

Sono state effettuate delle misure di temperatura su di un processore al variare del suo utilizzo



Poiché interessano le variazioni è possibile operare su dati modificati

% utilizzo CPU	Fattore di temperatura	Valori medi
0	3	1,333
0	1	
0	0	
20	4	5,333
20	7	
20	5	
50	9	8,333
50	6	
50	10	

$$\sum y_i = 45$$

$$\bar{y} = 5$$

$$CM = 5 \cdot 45 = 225$$

$$\sum y_i^2 = 317$$

$$Q_T = 317 - 225 = 92$$

L'effetto della temperatura si calcola come per la QT però considerando i valori medi di temperatura a condizioni costanti

$$Q_F = 3 \cdot 1,333^2 + 3 \cdot 5,333^2 + 3 \cdot 8,333^2 - 225 = 74$$

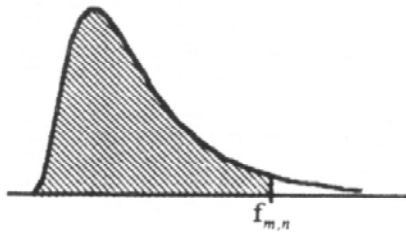
Variabilità imputabile all'errore $\longrightarrow Q_{ERR} = Q_T - Q_F = 92 - 74 = 18$



Causa	SS	GdL	MS	F _{sperim}	F _{teorica}
%utilizzo	74	2	37	12,33	5,143
ERR	18	6	3		
TOT	92	8			

$$0.95 = \mathbb{P}[F \leq f_{m,n}]$$

n: GdL dell'errore
m: GdL di fattore



	1	2	3	4	5	6	7	8	9	10
1	161,4	199,5	215,7	224,6	230,2	234	236,8	238,9	240,5	241,9
2	18,51	19	19,16	19,25	19,3	19,33	19,35	19,37	19,38	19,4
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,785
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964
5	6,608	5,786	5,409	5,192	5,05	4,95	4,876	4,818	4,772	4,735
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,06
7	5,591	4,737	4,347	4,12	3,972	3,866	3,787	3,726	3,677	3,637
8	5,318	4,459	4,066	3,838	3,688	3,581	3,5	3,438	3,388	3,347
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,23	3,179	3,137
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,02	2,978

$$F_{teorica} > F_{sperimentale} \longrightarrow$$

Non si può rifiutare l'ipotesi che non vi siano differenze significative nella misura della temperatura imputabili all'utilizzo della CPU



Analysis of Variance for CPU Temperature

Source	DF	SS	MS	F	P
% ut CPU	2	74,00	37,00	12,33	0,007
Error	6	18,00	3,00		
Total	8	92,00			

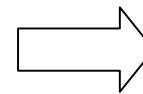
Output elaborato mediante software dedicato



Esempio 2: Due fattori controllati

Si valuta l'effetto sulla velocità di connessione del tipo di protocollo e del controllo di flusso usati

Protocollo	Controllo di flusso	velocità di trasferimento
A	A	3,911
A	B	3,811
A	C	3,871
B	A	3,921
B	B	3,812
B	C	3,888



Protocollo	Controllo di flusso	velocità di trasferimento
A	A	100
A	B	0
A	C	60
B	A	110
B	B	1
B	C	77

Protocollo	Controllo di flusso			somme	medie
	A	B	C		
A	100	0	60	160	53,33333
B	110	1	77	188	62,66667
somme	210	1	137	348	-
medie	105	0,5	68,5	-	58

$$\sum y_i = 348$$

$$\bar{y} = 58$$

$$CM = 58 \cdot 348 = 20184$$

$$\sum y_i^2 = 31630$$

$$Q_T = 31630 - 20184 = 11446$$



$$Q_R = 160 \cdot 53,33 + 188 \cdot 62,67 - 20184 = 130,67$$

Variazione per fattore riga

$$Q_C = 210 \cdot 105 + 1 \cdot 0,5 + 137 \cdot 68,5 - 20184 = 11251$$

Variazione per fattore colonna

$$GdL_{Totali} = n - 1$$

$$GdL_{Protocollo} = n_{Protocolli} - 1$$

$$GdL_{Controllo} = n_{Controlli} - 1$$

$$GdL_{Errore} = GdL_{Totali} - GdL_{Protocollo} - GdL_{Controllo}$$

Causa	GdL	SS	MS	F	P
Protocollo	1	130,7	130,7	4,06	0,181
Controllo	2	11251	5625,5	174,89	0,006
ERR	2	64,3	32,2		
TOT	5	11446			



Esempio 3: due fattori controllati con ripetute

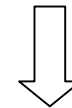
Protocollo	Controllo di flusso	velocità di trasferimento
A	A	3,911
A	A	3,913
A	A	3,916
A	B	3,811
A	B	3,812
A	B	3,808
A	C	3,871
A	C	3,864
A	C	3,867
B	A	3,921
B	A	3,902
B	A	3,915
B	B	3,812
B	B	3,82
B	B	3,818
B	C	3,888
B	C	3,89
B	C	3,889



Protocollo	Controllo di flusso	velocità di trasferimento
A	A	100
A	A	102
A	A	115
A	B	0
A	B	1
A	B	-3
A	C	60
A	C	53
A	C	56
B	A	110
B	A	91
B	A	104
B	B	1
B	B	9
B	B	7
B	C	77
B	C	79
B	C	78



Protocollo	Controllo di flusso		
	A	B	C
A	100,102,115	0,1,-3	60,53,56
B	110,91,104	1,9,7	77,79,80



Protocollo	Controllo di flusso			somme	medie
	A	B	C		
A	105,67	-0,67	56,33	161,33	53,78
B	101,67	5,67	78,67	186,00	62,00
somme	207,33	5,00	135,00	347,33	-
medie	103,67	2,50	67,50	-	57,89



$$SS_s = \sum_i \sum_j (y_{ij} \cdot \bar{y}_{ij}) - CM$$

Somma dei quadrati di sottotale

$$Q_{RC} = SS_s - Q_R - Q_C$$

Effetto di interazione tra i fattori

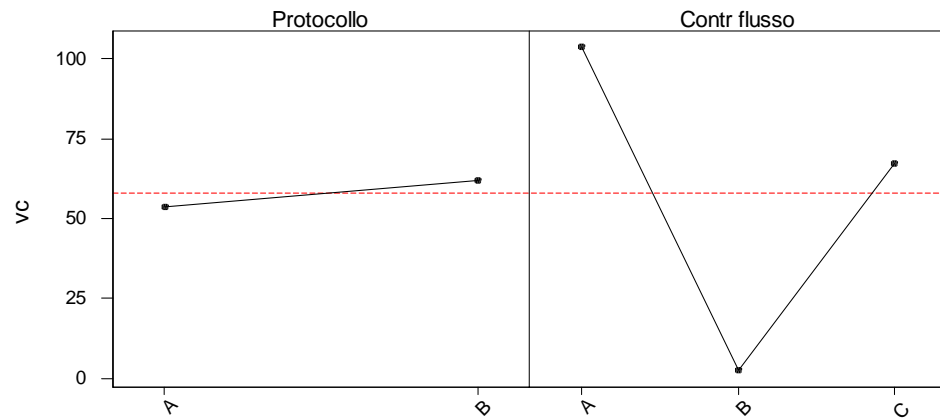
$$Q_{ERR} = Q_T - Q_R - Q_C - Q_{RC} = Q_T - SS_s$$

Causa	GdL	SS	MS	F	P
Protocollo	1	288	288	8,83	0,012
Controllo	2	31497,4	15748,7	482,93	0
Prot*Contr	2	500,3	250,2	7,67	0,007
ERR	12	391,3	32,6		
TOT	17	32677,1			

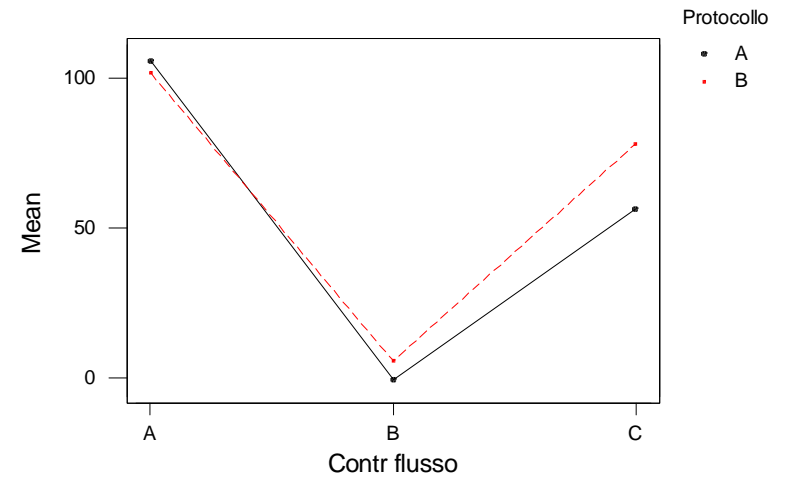
a=0,01



Main Effects Plot - Data Means for vc



Interaction Plot - Data Means for vc





Design Of Experiments

La realizzazione delle prove sperimentali rappresenta quasi sempre la principale fonte di costi sia come impiego di tempo che di risorse umane e materiali.

Mediante una ragionata applicazione delle tecniche del DOE è possibile ridurre drasticamente tali costi.

Principio di Pareto o "*legge 80/20*" (Vilfredo Pareto)

"The vital few and the trivial many" (Juran)

Prove pianificate avventatamente possono poi richiedere complicate tecniche di interpretazione dei dati sperimentali: in tal caso lo sperimentatore corre il rischio di perdere il contatto con la realtà del fenomeno oggetto di studio e fallire la comprensione dei risultati.



Da prove pianificate correttamente è possibile ottenere risultati ricchi di informazioni anche utilizzando semplici tecniche di analisi statistica.

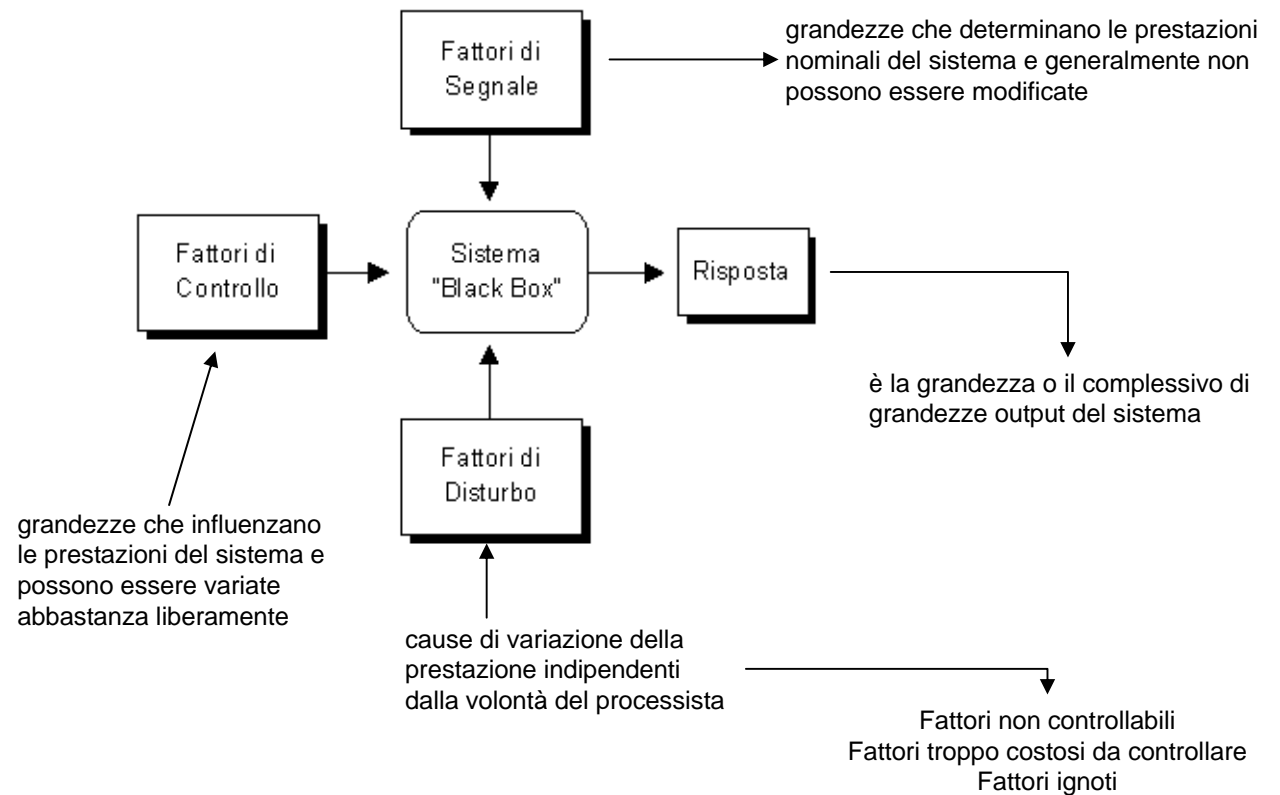


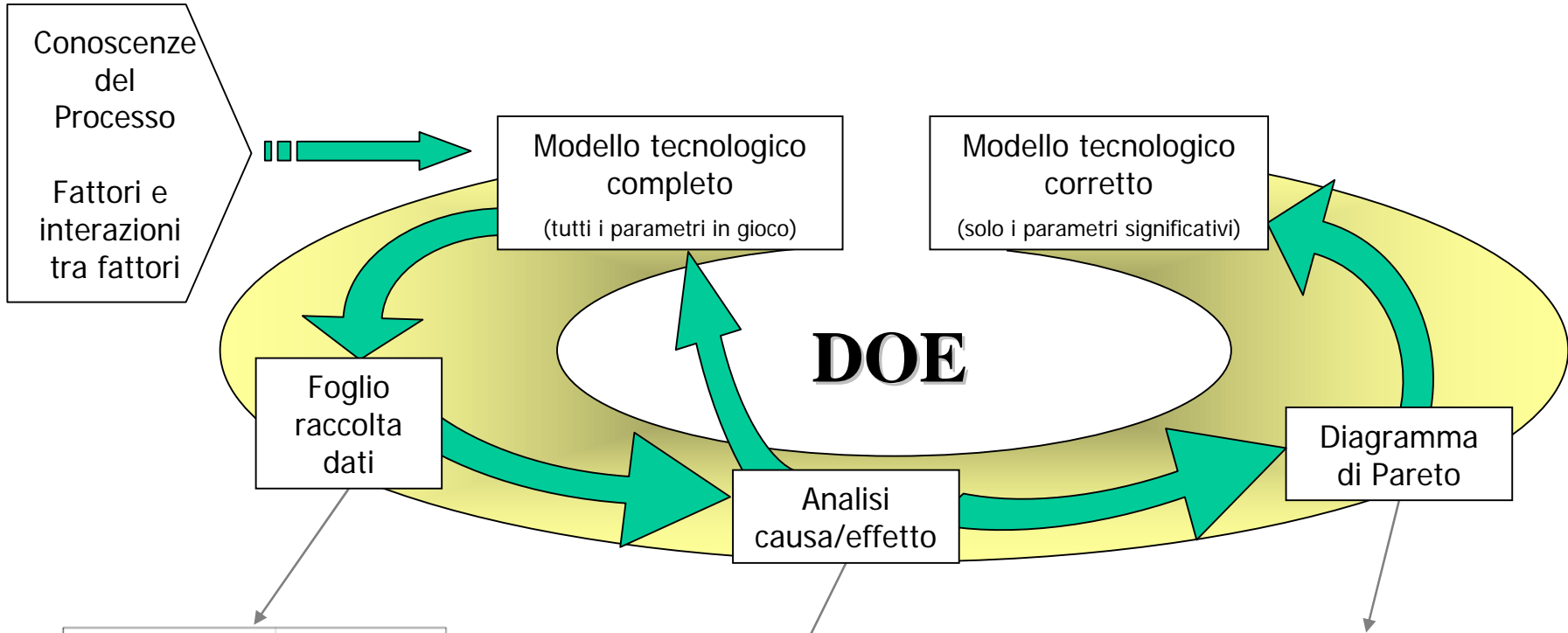
PROCESSO

Conoscenze:

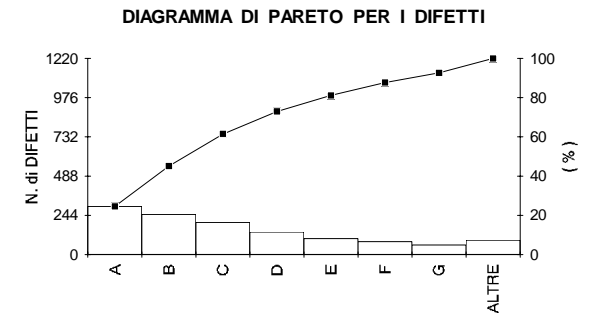
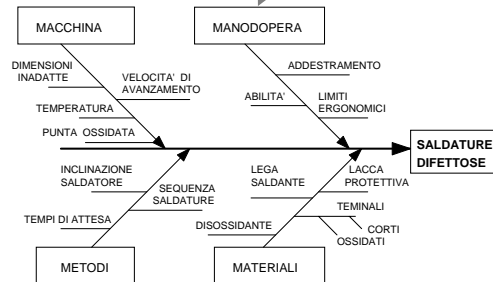
Schematizzazione:

- CONOSCENZA FISICA
- CONOSCENZA SPERIMENTALE
- CONOSCENZA INFORMALE



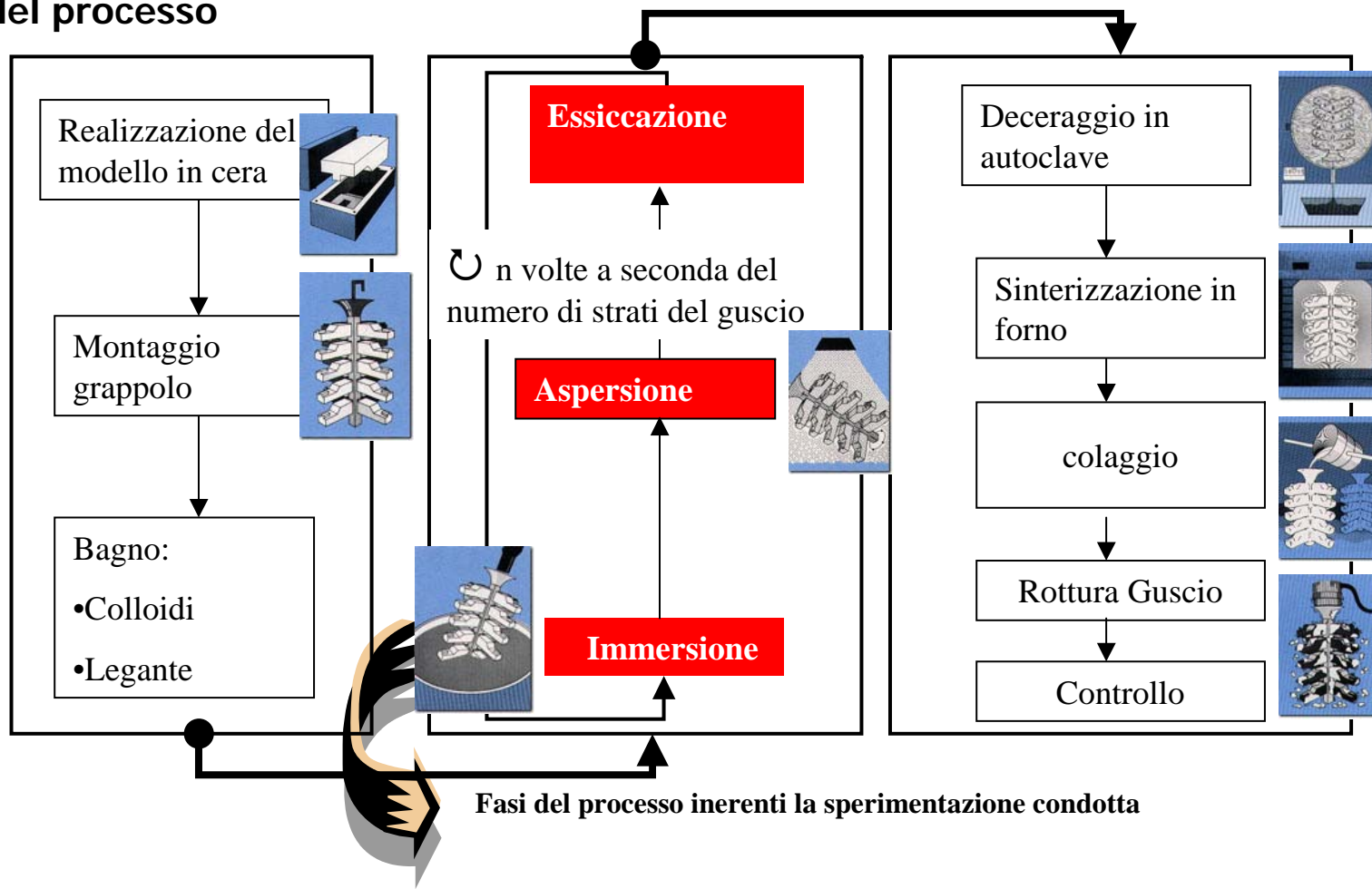


PRODOTTO: _____		DATA: _____										
TRATTAMENTO: _____		SECTORE: _____										
SPECIFICHE: _____		REPARTO: _____										
N° PEZZI ISPEZIONATI: _____		OPERATORE: _____										
N° TOTALE PEZZI: _____		TURNO: _____										
N° LOTTO: _____		NOTE: _____										
DIFETTI	DATA	07	07	107	117	127	157	167	177	187	197	TOT
Lampade tuminate		+++					+++	+++	+++		+++	40
Lampade mole serviate												6
Bastone fero difettosa			+++							+++	+++	28
Fero storto												9
Lampada sporca		+++			+++			+++		+++		52
Albi												11
TOTALE		20	10	9	13	4	14	21	15	15	25	145



Esempio: sperimentazione sul ciclo di fabbricazione delle forme ceramiche per il processo in cera persa

1: Analisi del processo

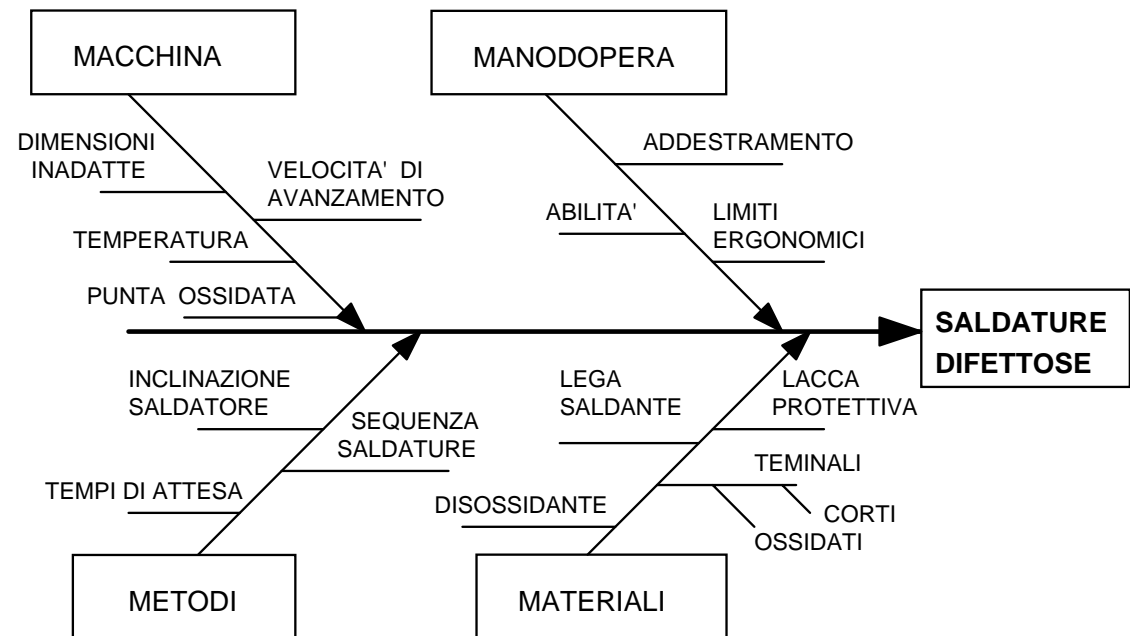




2: Digramma causa effetto

Diagramma a spina di pesce

Metodo 4M



A che cosa serve:

- A rappresentare tutte le cause possibili
- A ricercare le cause importanti

Come si applica:

- Analizza l'effetto rappresentato dalla barra più alta del diagramma di Pareto
- Le cause elencate devono essere molte ed analizzate da diversi punti di vista
- È uno strumento di gruppo che si applica con il metodo del brainstorming



3: Parametri in gioco e risposta del processo

Id.	Variabili
x_1	Velocità dell'aria (m/s)
x_2	Tempo intermedio d'essiccazione
x_3	Umidità relativa
x_4	T essiccazione primari
x_5	T essiccazione secondari

Risposta del processo

Permeabilità del guscio
Resistenza meccanica del guscio





4: Modello tecnologico lineare con interazioni multiple

$$y_{ijklmn} = \mu + \tau_i + \beta_j + \gamma_k + \nu_l + \eta_m + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\tau\nu)_{il} + (\tau\eta)_{im} + (\beta\gamma)_{jk} + (\beta\nu)_{jl} + (\beta\eta)_{jm} + (\gamma\eta)_{kl} + (\gamma\eta)_{km} + (\nu\eta)_{lm} + (\tau\beta\gamma)_{ijk} + (\tau\beta\nu)_{ijl} + (\tau\beta\eta)_{ijm} + (\tau\gamma\nu)_{ikl} + (\tau\gamma\eta)_{ikm} + (\tau\nu\eta)_{ilm} + (\beta\gamma\nu)_{jkl} + (\beta\gamma\eta)_{jkm} + (\beta\eta\nu)_{jml} + (\gamma\nu\nu)_{klm} + (\tau\beta\gamma\nu)_{ijkl} + (\tau\beta\gamma\nu)_{ijkm} + (\tau\beta\nu\eta)_{ijlm} + (\tau\gamma\nu\eta)_{iklm} + (\beta\gamma\nu\eta)_{jklm} + (\tau\beta\gamma\nu\eta)_{ijklm} + \varepsilon_{ijklmn} + \delta_n$$



5: Piano sperimentale delle prove

Piano fattoriale completo

5 fattori

2 livelli

6 ripetute

$N = 6 * 2^5 = 192$ prove

LE OSSERVAZIONI MANCANTI SONO INDICATE DA TRE ASTERISCHI *** (UNITA' DI MISURA: $[m^2 / (Pa * s^2)]$ I VALORI SONO IN SCALA $1 * 10^{11}$)

		Velocità (+)				Velocità (-)					
		Tempo (+)		Tempo (-)		Tempo (+)		Tempo (-)			
		Umidità (+)	Umidità (-)	Umidità (+)	Umidità (-)	Umidità (+)	Umidità (-)	Umidità (+)	Umidità (-)		
T strati primari (-)	T secondari (-)	183	166	226	158	252	135	327	166		
		177	171	232	170	256	135	***	181		
		197	166	227	160	220	145	317	186		
		172	162	220	152	248	137	298	169		
		169	166	217	168	222	134	294	163		
		169	172	***	***	222	128	293	164		
	T secondari (+)	181	169	181	192	203	165	201	199		
		175	179	175	194	188	164	235	199		
		181	185	164	193	211	165	***	192		
		181	172	166	194	200	156	237	205		
		175	183	173	187	201	153	259	213		
		179	171	175	196	188	172	2641	208		
		T strati primari (+)	T secondari (-)	195	206	222	147	184	208	209	244
				194	197	217	163	179	181	193	219
199	***			215	164	184	197	200	237		
183	187			215	159	166	197	191	234		
T secondari (+)	188		201	213	163	175	197	194	230		
	190		199	220	174	164	173	191	257		
	137		133	185	156	244	171	270	224		
	140		129	***	156	266	173	261	205		
T secondari (+)	158	***	187	158	267	166	244	239			
	142	***	169	157	204	166	255	210			
	156	132	164	146	206	170	224	206			
	144	***	184	148	228	159	234	188			

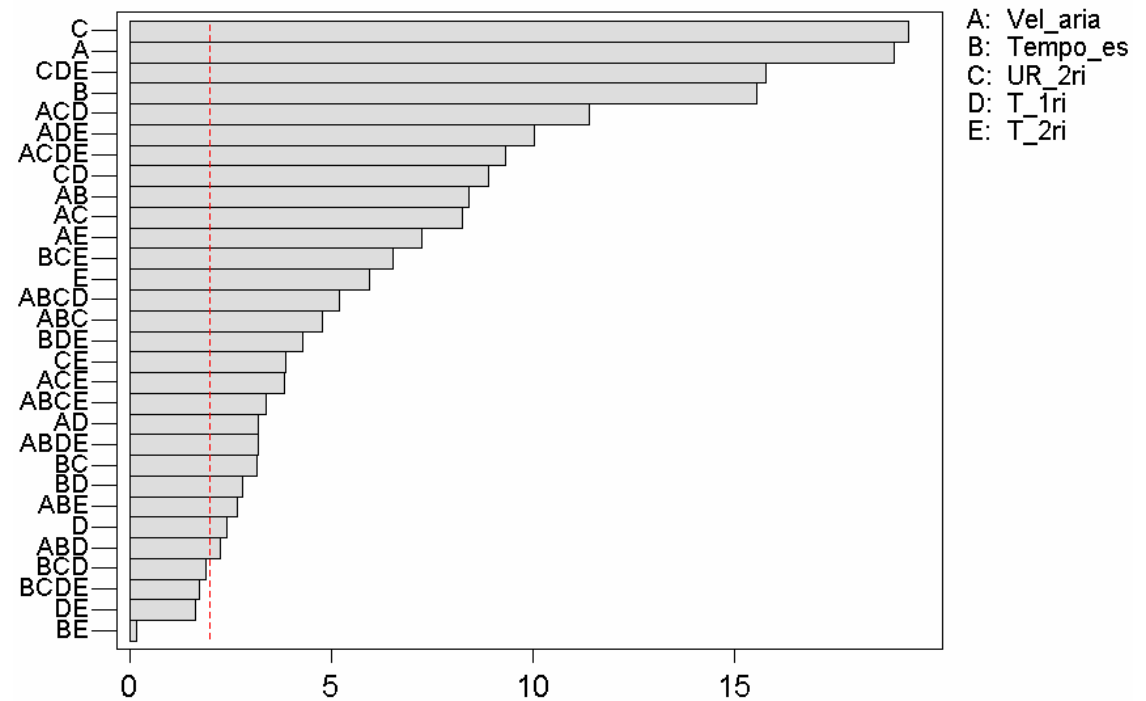


6: Diagramma di Pareto

eliminando il 20% dei difetti si elimina l'80% della difettosità

il 20% dei dipendenti fa l'80% delle assenze

Il 20% della popolazione possiede l'80% delle ricchezze





7: Modello tecnologico ridotto ai parametri significativi

$$y_{ijklmn} = \mu + \tau_i + \beta_j + \gamma_k + \nu_l + \eta_m + (\tau\beta)_{ij} + (\beta\gamma)_{jk} + (\beta\nu)_{jl} + (\beta\eta)_{jm} + (\gamma\eta)_{km} + (\nu\eta)_{lm} + (\tau\beta\gamma)_{ijk} + (\tau\beta\eta)_{ijm} + (\tau\gamma\eta)_{ikm} + (\tau\nu\eta)_{ilm} + (\beta\gamma\nu)_{jkl} + (\beta\gamma\eta)_{jkm} + (\gamma\nu\nu)_{klm} + (\tau\beta\gamma\nu)_{ijkl} + (\tau\beta\nu\eta)_{ijlm} + (\beta\gamma\nu\eta)_{jklm} + \varepsilon_{ijklmn} + \delta_n$$



8: Nuova sperimentazione fattoriale su più livelli e meno parametri

Dal diagramma di Pareto si vede che i parametri più significativi sono (nell'ordine)

C
A
B

quindi:

Piano fattoriale completo

3 fattori (A,B,C)

3 livelli (alto medio basso)

4 ripetute

$N = 4 * 3^3 = 108$ prove

(NB: molte prove sono già state effettuate)



Piano fattoriale frazionario

È possibile impiegare una o più colonne di interazione per stimare un effetto principale

k numero di variabili

—————> 2^{k-p} numero di prove nel piano frazionalizzato

p variabili assegnate alle colonne di interazione

Esempio: 4 fattori 2 livelli

N° condizione di prova	X ₁	X ₂	X ₃	X ₄
1	+	+	+	+
2	+	+	+	-
3	+	+	-	+
4	+	+	-	-
5	+	-	+	+
6	+	-	+	-
7	+	-	-	+
8	+	-	-	-
9	-	+	+	+
10	-	+	+	-
11	-	+	-	+
12	-	+	-	-
13	-	-	+	+
14	-	-	+	-
15	-	-	-	+
16	-	-	-	-

Piano fattoriale completo
 $2^4 = 16$ condizioni di prova

Si utilizza l'interazione $1*2*3$
per stimare l'effetto principale 4

N° condizione di prova	Media	1	2	3	4	1*2	1*3	1*4	2*3	2*4	3*4	1*2*3	1*2*4	1*3*4	2*3*4	1*2*3*4
1	+	-	-	-	+	+	+	+	+	+	-	-	-	-	-	+
2	+	+	-	-	+	-	-	+	+	-	-	+	-	-	-	+
3	+	-	+	-	+	-	+	-	-	+	-	+	-	+	-	+
4	+	+	+	-	+	-	-	-	-	+	-	-	+	+	+	+
5	+	-	-	+	+	+	-	-	-	+	+	+	-	-	-	+
6	+	+	-	+	-	-	+	-	-	+	-	-	+	-	+	+
7	+	-	+	+	-	-	-	+	+	-	-	-	+	+	-	+
8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Piano fattoriale frazionario
 $2^{4-1} = 8$ condizioni di prova



Esempi di intervalli di confidenza e relative Analisi della varianza

Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Rtm
All Pairwise Comparisons among Levels of m

m = 1 subtracted from:

m	Lower	Center	Upper
2	-4,438	-1,849	0,7396
3	-4,163	-1,574	1,0146
4	-2,954	-0,144	2,6650

-----+-----+-----+-----
 (------*-----)
 (------*-----)
 (------*-----)
 -----+-----+-----+-----
 -2,5 0,0 2,5

m = 2 subtracted from:

m	Lower	Center	Upper
3	-2,314	0,2750	2,864
4	-1,105	1,7047	4,514

-----+-----+-----+-----
 (------*-----)
 (------*-----)
 -----+-----+-----+-----
 -2,5 0,0 2,5

m = 3 subtracted from:

m	Lower	Center	Upper
4	-1,380	1,430	4,239

-----+-----+-----+-----
 (------*-----)
 -----+-----+-----+-----
 -2,5 0,0 2,5

Tukey Simultaneous Tests
Response Variable Rtm
All Pairwise Comparisons among Levels of m

m = 1 subtracted from:

Level	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
2	-1,849	0,9976	-1,854	0,2527
3	-1,574	0,9976	-1,578	0,3943
4	-0,144	1,0826	-0,133	0,9991

m = 2 subtracted from:

Level	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
3	0,2750	0,9976	0,2757	0,9927
4	1,7047	1,0826	1,5747	0,3962

m = 3 subtracted from:

Level	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
4	1,430	1,083	1,321	0,5513

Tukey 95.0% Simultaneous Confidence Intervals
Response Variable Rv
All Pairwise Comparisons among Levels of media

media = 1 subtracted from:

media	Lower	Center	Upper
2	-6,702	-3,932	-1,161
3	3,250	6,021	8,791

---+-----+-----+-----+-----
 (----*---)
 (----*----)
 ---+-----+-----+-----+-----
 -6,0 0,0 6,0 12,0

media = 2 subtracted from:

media	Lower	Center	Upper
3	7,182	9,952	12,72

---+-----+-----+-----+-----
 (----*---)
 ---+-----+-----+-----+-----
 -6,0 0,0 6,0 12,0

Tukey Simultaneous Tests
Response Variable Rv
All Pairwise Comparisons among Levels of media

media = 1 subtracted from:

Level	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
2	-3,932	1,169	-3,362	0,0028
3	6,021	1,169	5,148	0,0000

media = 2 subtracted from:

Level	Difference of Means	SE of Difference	T-Value	Adjusted P-Value
3	9,952	1,169	8,510	0,0000