Progetto e miglioramento del processo produttivo

Introduzione

Linee guida della progettazione

Analisi della varianza

Introduzione

L’introduzione formale di una metodologia di **programmazione degli esperimenti** nei primi stadi del ciclo di sviluppo del prodotto e del processo produttivo è spesso molto rilevante nel successo complessivo del prodotto, in termini di progettazione di nuovi prodotti, di miglioramento del progetto di prodotti esistenti e di ottimizzazione del processo produttivo.

L’uso efficace di una solida metodologia di programmazione statistica degli esperimenti può portare a prodotti caratterizzati da:

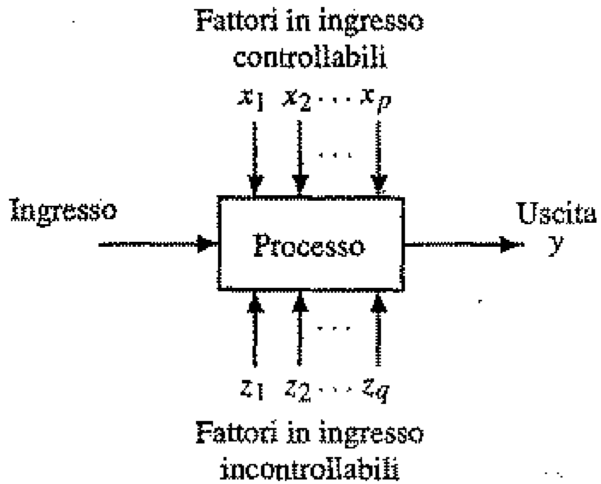
- maggiore facilità di realizzazione;
- migliore affidabilità;
- migliori prestazioni sul campo.

Inoltre la programmazione degli esperimenti può anche migliorare molto le attività di sviluppo del processo produttivo e di messa a punto.

Un esperimento programmato è una prova, o una serie di prove, in cui vengono fatte variare deliberatamente le variabili di ingresso di un processo, in modo da poter osservare e identificare le corrispondenti variazioni della risposta in uscita.

Un processo può essere visualizzato come un complesso di macchine, metodi e persone che trasformano il materiale in ingresso nel prodotto in uscita.

Questo prodotto in uscita ha una o più caratteristiche di qualità o risposte osservabili. Alcune delle variabili di processo, siano x_1, x_2,... , x_p, sono **controllabili,** mentre altre, siano z_1, z_2, …, z_q, sono **incontrollabili** (anche se queste ultime potrebbero essere controllabili nel corso dell'esperimento). A volte questi fattori incontrollabili sono detti **fattori di rumore.**



Gli obiettivi dell'esperimento possono comprendere:

- La determinazione di quali variabili hanno più influenza sulla risposta y;
- La determinazione di come aggiustare le *x* più influenti in modo che la risposta *y* sia vicina alla richiesta della specifica;
- La determinazione di come aggiustare le x più influenti in modo che la variabilità di y sia piccola;
- La determinazione di come aggiustare le x influenti in modo che l'effetto delle variazioni non controllabili z sia minimizzato.

I metodi della programmazione degli esperimenti possono essere usati sia nello sviluppo sia nella messa a punto del processo per migliorarne le prestazioni o per ottenere un processo **robusto** ovvero **insensibile** alle sorgenti esterne di variabilità.

I metodi di controllo statistico di processo (SPC) e la programmazione degli esperimenti, entrambi strumenti potenti per migliorare ed ottimizzare il processo, sono tra loro collegati:

SPC → metodo statistico **passivo** → osserviamo il processo in attesa di certe informazioni che potranno condurre a cambiamenti utili

↓

se il processo è sotto controllo, l'osservazione passiva non può produrre molta informazione utile

Programmazione degli esperimenti → metodo statistico **attivo** → informazione che può condurre ad un miglioramento del processo

- E' necessaria una buona conoscenza tecnica nella scelta dei parametri da investigare;
- La scelta dei fattori del processo dovrebbe essere fatta coinvolgendo gli attori con più esperienza, con maggiore conoscenza del processo;
- Per ogni fattore bisogna decidere il numero di livelli da considerare. Il numero minimo di livelli è due, tipicamente basso e alto (spesso indicati con 1 e 2 oppure –1 e +1);
- Per ogni fattore, per ogni livello, si possono scegliere più osservazioni, dette replicazioni o ripetizioni (replication);
- La ripetizione degli esperimenti consente di ottenere una stima dell'errore sperimentale. Tale quantità è importante per determinare in quale misura le differenze osservate fra i dati sono effettivamente delle differenze statistiche;
- L'obiettivo è di minimizzare il numero delle prove, per ovvi motivi di costo.

Il numero delle prove necessarie si calcola come:

N° prove = n° ripetizioni x (n° livelli)^{nf} dove nf = numero di fattori

Linee guida per la progettazione degli esperimenti

1. Riconoscimento e definizione del problema

Una chiara enunciazione del problema e degli obiettivi dell'esperimento spesso contribuisce in modo sostanziale ad una migliore comprensione del processo e alla successiva soluzione del problema. E' importante sollecitare contributi da tutte le parti coinvolte.

2. Scelta dei fattori e dei livelli

Occorre una determinata esperienza per scegliere quali fattori far variare nell'esperimento, l'intervallo in cui ogni fattore deve essere fatto variare, e gli specifici livelli ai quali devono essere eseguite le prove. Quando l'obiettivo è l'individuazione dei fattori o la caratterizzazione del processo è di solito meglio tenere basso il numero di livelli dei fattori (il più delle volte se ne utilizzano due).

3. Scelta della variabile di output

Va scelta con oculatezza: deve rappresentare bene l'oggetto di studio, la capacità di misura deve essere adeguata. Risposte multiple non sono rare.

Tali 3 fasi sono spesso racchiuse nel termine di <u>programmazione</u> <u>pre-sperimentale</u>

4. Scelta del piano sperimentale

La scelta del piano comporta la considerazione della dimensione del campione (numero di ripetizioni), la scelta di un opportuno ordine delle prove per l'esecuzione dell'esperimento,e se intervengono bloccaggi o altre restrizioni alla casualizzazione.

5. Esecuzione dell'esperimento

Si intuisce l'importanza di eseguire il piano di prove secondo condizioni corrette.

6. Analisi dei dati

L'applicazione di metodi statistici permette di ottenere indicazioni obiettive e non basate su giudizi personali o esperienze passate. Permettono di mettere in luce le caratteristiche del processo con un determinato livello di confidenza.

7. Conclusioni

Dopo l'analisi dei dati, si propone una linea di azione focalizzata. L'azione correttiva è dettata dai risultati dei metodi statistici. Sono molto utili anche delle rappresentazioni grafiche, soprattutto nel caso di presentazione dei dati a terzi.

NB: la sperimentazione è <u>iterativa,</u> per cui man mano che un programma di sperimentazione procede, spesso cambiano i fattori, i loro livelli e le variabili di risposta studiate.

L'analisi della varianza (ANOVA)

ANOVA
ANalysis Of VAriance → serve ad ottenere informazioni sugli effetti esercitati su di una variabile di nostro interesse da determinati fattori

Consente sostanzialmente di ripartire la varianza sperimentale in aliquote s-indipendenti → al fine di isolare quelle imputabili singolarmente ai fattori per il cui studio è stato formulato il piano sperimentale

Varianza totale = la somma di queste aliquote + varianza residua

→ costituisce una misura della variabilità fisiologica che caratterizza lo specifico contesto sperimentale

NB: affinché un fattore determini un effetto significativo (o, come si dice, sia esso stesso significativo) deve produrre un'aliquota della varianza maggiore di quella residua.

Analisi della varianza ad una via

Si considera il caso ad un solo fattore con a diversi livelli disponendo di n osservazioni sperimentali per ciascun livello.

Ogni osservazione sperimentale y_{ij} può essere vista come la somma delle seguenti componenti:

$$\boxed{y_{ij} = \mu + \tau_i + \varepsilon_{ij}} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

overall mean → è la media generale comune a tutte le osservazioni → è il risultato che si avrebbe in assenza di fluttuazioni

factor effect → l'effetto imputabile al fattore in esame al livello i-esimo

random error component → è una componente di errore casuale relativa all'osservazione j-esima in corrispondenza del livello i-esimo del fattore

$$E\{\varepsilon_{ij}\} = 0$$

$$\mathrm{var}\{\varepsilon_{ij}\} = \sigma^2$$

Ipotesi su cui è fondato il modello di analisi:

- dipendenza lineare degli effetti dei fattori;
- additività degli effetti dei fattori;
- le osservazioni devono essere fatte in ordine casuale, in modo tale che l'ambiente in cui i fattori sono utilizzati (spesso si parla anche di unità sperimentali) sia il più uniforme possibile;

  Piano sperimentale completamente casualizzato

- gli errori ε_{ij} sono variabili casuali indipendenti e di distribuzione normale;

  $$\varepsilon_{ij} = N(0;\sigma^2)$$

- varianza σ^2 costante per tutti i livelli del fattore.

Scelta dei livelli del fattore a:

- Modello ad effetti fissi (livelli scelti esplicitamente dallo sperimentatore);
- Modello ad effetti casuali o a componenti di varianza (estratti come un campione casuale).

Il Modello ad effetti fissi

Gli effetti dei fattori sono definiti come deviazioni dalla media generale: $\sum_{i=1}^{a} \tau_i = 0$

Si definiscono le seguenti grandezze :

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{i.} = y_{i.} / n \qquad i = 1,2,\ldots\ldots,a$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \bar{y}_{..} = y_{..} / N \qquad N = an$$

Livello del fattore	Osservazioni	Totali	Medie
1	$y_{11}, y_{12}, \ldots\ldots. y_{1n}$	$y_{1.}$	$\bar{y}_{1.}$
2	$y_{21}, y_{22}, \ldots\ldots. y_{2n}$	$y_{2.}$	$\bar{y}_{2.}$
....	....	....	....
a	$y_{a1}, y_{a2}, \ldots\ldots. y_{an}$	$y_{a.}$	$\bar{y}_{a.}$
		$y_{..}$	$\bar{y}_{..}$

Test d'ipotesi

Verificare l'efficacia del fattore in esame significa verificare l'ipotesi nulla H_0:

$$H_0 : \{\tau_i = 0, i = 1,\dots,a\}$$

mentre l'ipotesi contraria H_1 è

$$H_1 : \{\tau_i \neq 0, \text{per almeno un livello i}\}$$

Per verificare le ipotesi del test, si effettua una procedura di verifica, che si basa sulla ripartizione della variabilità totale dei dati, in modo da isolare l'aliquota imputabile al fattore in esame.

Si considera la somma generale corretta dei quadrati definita:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{..})^2 \longleftarrow \text{misura della variabilità totale}$$

Tale somma può essere riscritta come:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij} - \overline{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}[(y_{ij} - \overline{y}_{i.}) + (\overline{y}_{i.} - \overline{y}_{..})]^2$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n}[(y_{ij}-\bar{y}_{i.})^2 + 2\sum_{i=1}^{a}\sum_{j=1}^{n}[(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..})$$

$$\sum_{i=1}^{a}\sum_{j=1}^{n}[(y_{ij}-\bar{y}_{i.})(\bar{y}_{i.}-\bar{y}_{..})] = \sum_{i=1}^{a}\left[(\bar{y}_{i.}-\bar{y}_{..})\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})\right] =$$

$$= \sum_{i=1}^{a}\left[(\bar{y}_{i.}-\bar{y}_{..})\left(\sum_{j=1}^{n}(y_{ij}-n\sum_{j=1}^{n}y_{ij}/n)\right)\right] = 0$$

$$\boxed{\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2 + n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2}$$

$$SS_{Totale} = SS_{Errore} + SS_{Fattore}$$

la somma dei quadrati degli scarti rispetto alla media generale è ripartibile in due aliquote indipendenti costituite rispettivamente dalla somma degli scarti quadratici rispetto alle medie parziali e da n volte la somma degli scarti quadratici delle a medie parziali rispetto a quella globale

I gradi di libertà corrispondono al numero di elementi che risultano realmente indipendenti, atteso l'impiego di relazioni che legano il loro insieme.

$SS_{Fattore} \longrightarrow$ a-1 g.d.l.
$SS_{Errore} \longrightarrow$ a(n-1) g.d.l.
$SS_{Totale} \longrightarrow$ an-1 g.d.l.

Il rapporto tra una somma dei quadrati ed il suo numero di gradi di libertà si chiama <u>media dei quadrati.</u>

$$MS_{Errore} = \frac{SS_{Errore}}{a(n-1)} \qquad\qquad E(MS_{Errore}) = \sigma^2$$

$$MS_{Fattore} = \frac{SS_{Fattore}}{a-1} \quad\longrightarrow\quad E(MS_{Fattore}) = \sigma^2 + n\frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$$

Se H_0 è vera $\longrightarrow$ $MS_{Fattore}$ è uno stimatore non distorto di σ^2;
Se H_1 è vera $\longrightarrow$ $MS_{Fattore}$ <u>non</u> è uno stimatore corretto di σ^2.

Se l'ipotesi nulla H_0 è vera, il rapporto:

$$\frac{SS_{Fattore}/a-1}{SS_{Errore}/a(n-1)} = \frac{MS_{Fattore}}{MS_{Errore}} \approx F_{a-1,a(n-1)}$$

costituisce una variabile aleatoria F, detta di Fisher, con a-1 e a(n-1) gradi di libertà.

La verifica di ipotesi per l'eguaglianza delle medie sui livelli del fattore si basa sul valore assunto dalla statistica test F_0:

$$F_0 = \frac{MS_{Fattore}}{MS_{Errore}}$$

$F_0 > F_{\alpha,a-1,a(n-1)} \longrightarrow$ Le medie ai livelli del fattore sono differenti; quindi l'effetto del fattore è significativo: rifiuto l'ipotesi nulla H_0.

Ovvero H_0 è rigettata se al valore osservato F_0, è associata una coda di probabilità:

$$p = \Pr\{F_{(a-1),a(n-1)} > F_0\} < \alpha = \Pr\{\Im\}$$

essendo α la probabilità che definisce la zona di rigetto $\Im$

Se la varianza imputabile al fattore in esame è significativamente più grande della varianza dell'errore sperimentale dobbiamo ritenere che ciò che abbiamo osservato è verosimilmente una conseguenza di una realtà non conforme ad H_0.

Origine della varianza	Somma dei quadrati degli scarti	Gradi di libertà	Scarto quadratico medio (s.q.m.)	Valore atteso dello s.q.m. $E\{s.q.m.\}$
Fattore (var. tra classi)	$n\sum_{i=1}^{a}(\bar{y}_{i.}-\bar{y}_{..})^2$	$a-1$	$\frac{n\sum_f}{(a-1)}$	$\sigma^2+\frac{n}{a-1}\sum_{i=1}^{a}\tau_i^2$
Errore (var. interclassi)	$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$	$a(n-1)$	$\frac{\sum_e}{a(n-1)}$	σ^2
Totale	$\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$	$an-1$	$\frac{\sum_t}{an-1}$	

Relativamente alle medie dei quadrati:

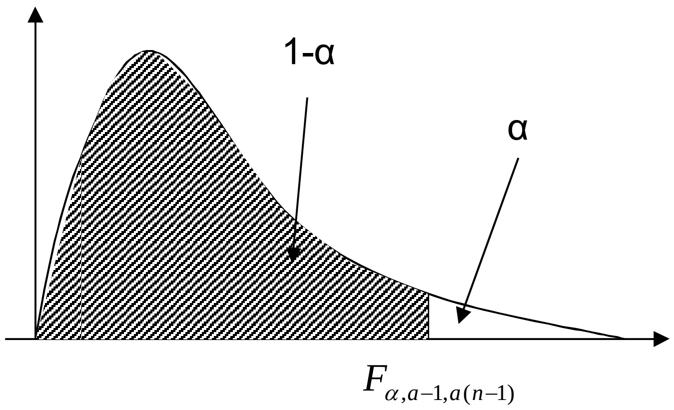
$SS_{Errore}/[a(n-1)]$ è <u>sempre</u> uno stimatore corretto di σ^2

$SS_{Fattore}/(a-1)$ è uno stimatore corretto di σ^2 <u>solo</u> se il fattore non ha effetto

Se il fattore è <u>significativo</u>, la media dei quadrati relativa al fattore tende a sovrastimare il valore di σ^2, per cui ci si aspetta in questo caso un valore del rapporto:

$$F_0 = \frac{SS_{Fattore}/(a-1)}{SS_{Errore}/[a(n-1)]} > 1$$

quindi la regione critica è costituita dalla sola coda <u>superiore</u> della distribuzione F.



Riassumendo: $SS_t = SS_e + SS_f$

SS_t *sum of squares corrected for the mean*

SS_e *sum of squares within groups*

SS_f *sum of squares between groups*

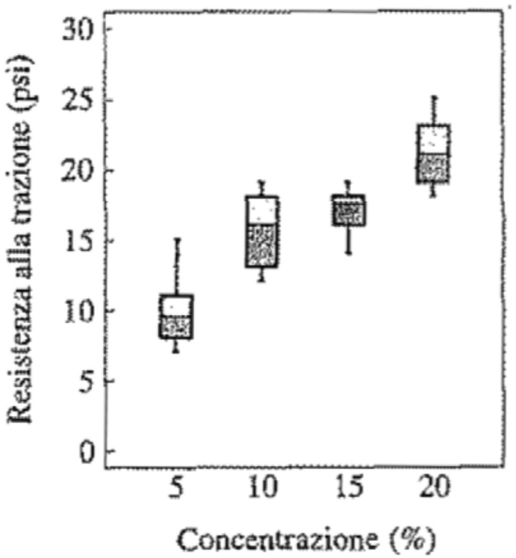
$$MS_{errore} = \frac{SS_{errore}}{a(n-1)}$$ *mean squares within groups*

$$MS_{fattore} = \frac{SS_{fattore}}{a-1}$$ *mean squares between groups*

$$F_0 = \frac{MS_{fattore}}{MS_{errore}}$$

Esempio: Un fabbricante di carta utilizzata per la produzione di sacchetti per la spesa è interessato al miglioramento della resistenza alla trazione del prodotto. Le specifiche del processo produttivo normalmente richiedono una concentrazione del 10% di legno duro nella pasta; con questo livello la carta ha una resistenza media alla trazione di circa 15 psi. Il processo è in stato di controllo statistico.
L'ingegnere di processo e gli operatori sospettano che la resistenza alla trazione sia una funzione della concentrazione di legno duro nella pasta e che la resistenza alla trazione sia maggiore ad una maggiore concentrazione di legno duro. L'economia del processo impone che il campo di variazione di interesse pratico per la concentrazione di legno duro sia tra 5% e 20%. L'ingegnere di processo decide di studiare quattro livelli di concentrazione di legno duro: 5%, 10%, 15% e 20%. Egli decide inoltre di fabbricare sei campioni di prova per ogni livello di concentrazione, usando un impianto pilota. Tutti i 24 campioni sono valutati con uno strumento di laboratorio per la resistenza alla trazione in ordine casuale.

Concentrazione di legno duro (%)	Osservazioni 1	2	3	4	5	6	Totali	Medie
5	7	8	15	11	9	10	60	10.00
10	12	17	13	18	19	15	94	15.67
15	14	18	19	17	16	18	102	17.00
20	19	25	22	23	18	20	127	21.17
							383	15.96



$$SS_{\text{Totale}} = \sum_{i=1}^{4}\sum_{j=1}^{6} y_{ij}^2 - \frac{y_{..}^2}{an}$$
$$= (7)^2 + (8)^2 + \cdots + (20)^2 - \frac{(383)^2}{24} = 512.96$$
$$SS_{\text{Fattore}} = \sum_{i=1}^{4} \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an}$$
$$= \frac{(60)^2 + (94)^2 + (102)^2 + (127)^2}{6} - \frac{(383)^2}{24} = 382.79$$
$$SS_{\text{Errore}} = SS_{\text{Totale}} - SS_{\text{Fattore}}$$
$$= 512.96 - 382.79 = 130.17$$

$$F_0 > F_{0.01,3,20} = 4{,}94$$

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Media dei quadrati	F_0	P
Concentrazione di legno duro	382.79	3	127.60	$F_0 = 19.61$	3.59×10^{-6}
Errore	130.17	20	6.51		
Totale	512.96	23			

Confronto tra le medie individuali

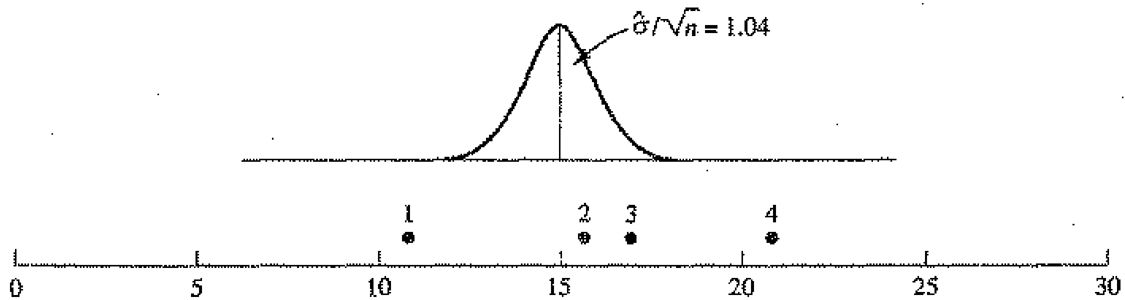
Se H_0 è valida ⟶ $\tau_i = 0 \quad \forall i = \{1, 2, ..., a\}$

quindi le medie $\bar{y}_{i.}$ dovrebbero comportarsi come un insieme di osservazioni estratte a caso dalla stessa distribuzione normale con media μ e deviazione standard $\sigma/n^{1/2}$.

σ incognito ⟶ stimo la σ come: $\hat{\sigma} = \sqrt{MS_{Errore}}$

utilizzo allora la distribuzione T di Student con fattore di scala pari a $(MS_{Errore}/n)^{1/2}$

se non c'è modo di sistemare la distribuzione in modo tale che le medie si presentino come tipiche estrazioni casuali ⟶ l'effetto del fattore è significativo



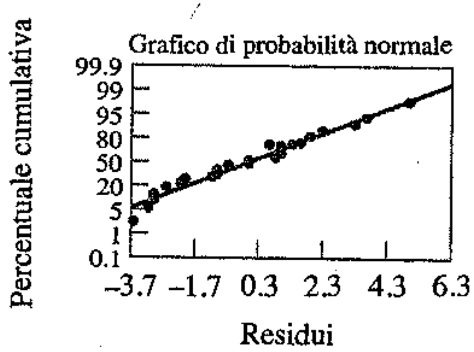
Analisi dei residui

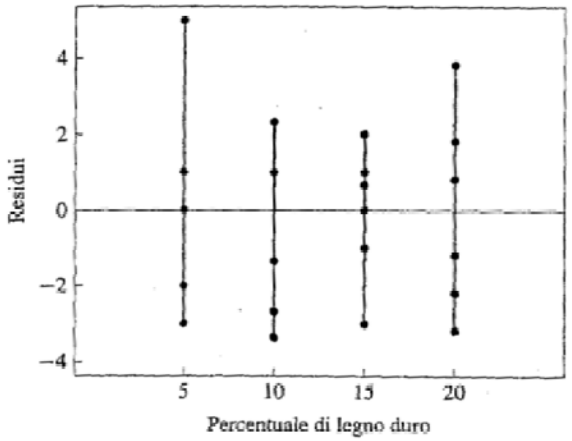
Si verifica l'assunzione che gli errori del modello siano normalmente ed indipendentemente distribuiti, con la stessa varianza per ogni livello del fattore.

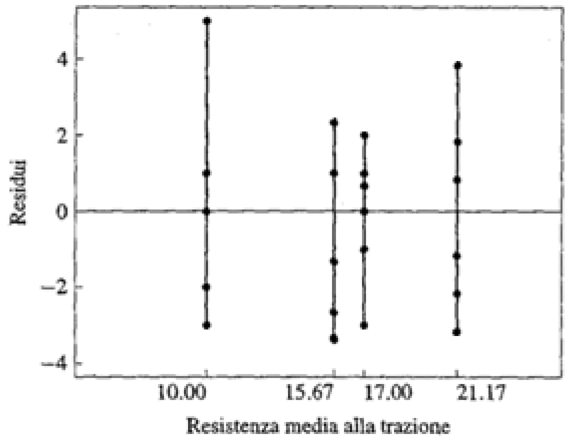
$$e_{ij} = y_{ij} - \bar{y}_{i.} \longrightarrow \text{assunzione di normalità}$$

$\downarrow$

assunzione di uguaglianza delle varianze







Il Modello a componenti di varianza

$$\boxed{y_{ij} = \mu + \tau_i + \varepsilon_{ij}} \qquad \begin{cases} i = 1,2,\ldots,a \\ j = 1,2,\ldots,n \end{cases}$$

Variabili casuali indipendenti ⟶ Altre ipotesi del modello:
ε_{ij} indipendenti e $N(0,\sigma^2)$;
τ_i indipendenti e $N(0,\sigma_\tau^2)$

$$\sigma_y^{\ 2} = \sigma_t^{\ 2} + \sigma^2 \qquad SS_{Totale} = SS_{Errore} + SS_{Fattore}$$

$$F_0 = \frac{MS_{Fattore}}{MS_{Errore}} \longrightarrow \begin{array}{l} H_0 : \sigma_\tau^2 = 0 \\ H_1 : \sigma_\tau^2 > 0 \end{array}$$

Se $\sigma_\tau^2 = 0$ ⟶ τ_i identici

Se $\sigma_\tau^2 > 0$ ⟶ effetto del livello del fattore presente

Stime delle componenti di varianza ⟶ $\hat{\sigma}^2 = MS_{Errore}$ $\qquad \hat{\sigma}_\tau^{\ 2} = \dfrac{MS_{Fattore} - MS_{Errore}}{n}$

Il piano a blocchi casualizzati

In molti problemi sperimentali è necessario programmare l'esperimento in modo che la variabilità proveniente da variabili di disturbo possa essere controllata. Questo tipo di piano sperimentale è detto **piano a blocchi completi e casualizzato**.

Supponiamo di essere interessati ad un solo fattore con *a* livelli e che l'esperimento sia condotto in *b* blocchi. Le osservazioni possono essere rappresentate con il modello:

$$y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij} \qquad \begin{cases} i = 1,2\ldots,a \\ j = 1,2,\ldots,b \end{cases}$$

μ è la media generale

τ_i è l'effetto dell'i-esimo livello del fattore $\rightarrow \sum_{i=1}^{a} \tau_i = 0$

β_j è l'effetto del j-esimo blocco $\rightarrow \sum_{j=1}^{b} \beta_j = 0$

ε_{ij} è l'errore indipendente con media nulla e varianza σ^2

Siamo interessati a testare l'eguaglianza degli effetti dei fattori, ossia le ipotesi:

$$H_0 \quad : \quad \tau_i = 0, i = 1, \dots, a$$

$$H_1 \quad : \quad \tau_i \neq 0 \, per \; almeno \; una \; i$$

Somma generale: $$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 = \sum_{i=1}^{a}\sum_{j=1}^{b}[(\bar{y}_{i.} - \bar{y}_{..}) + (\bar{y}_{.j} - \bar{y}_{..})(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})]^2$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{..})^2 = b\sum_{i=1}^{a}(\bar{y}_{i.} - \bar{y}_{..})^2 + a\sum_{j=1}^{b}(\bar{y}_{.j} - \bar{y}_{..})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}(y_{ij} - \bar{y}_{.j} - \bar{y}_{i.} + \bar{y}_{..})^2$$

$$SS_{totale} = SS_{fattore} + SS_{blocchi} + SS_{errore}$$

Gradi di libertà: $$ab - 1 = (a-1) + (b-1) + (a-1)(b-1)$$

Test: $$F_0 = \frac{MS_{Fattore}}{MS_{Errore}}$$

Analisi della varianza a due o più vie

L'analisi della varianza può essere estesa per trattare gli esperimenti fattoriali a due o più fattori.

Indichiamo i due fattori con A e B, con a livelli per il fattore A e b livelli per il fattore B, con n replicazioni di ciascuna prova; le abn prove dovrebbero essere eseguite in ordine casuale.

y_{ijk} osservazione nella cella ij-esima nella k-esima replicazione

Le osservazioni ottenute da un esperimento fattoriale a due fattori possono essere descritte dal modello:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk} \qquad \begin{cases} i = 1,2,\ldots.., a \\ j = 1,2,\ldots., b \\ k = 1,2,\ldots., n \end{cases}$$

μ è la media generale

τ_i è l'effetto dell'i-esimo livello del fattore A

β_j è l'effetto del j-esimo livello del fattore B

$(\tau\beta)_{ij}$ è l'effetto dell'interazione fra A e B

ε_{ijk} è l'errore casuale normale indipendente con media nulla e varianza σ^2

Siamo interessati ad un test dell'ipotesi che l'effetto del fattore A sia non significativo, o che l'effetto del fattore B sia non significativo, o che non sia significativa l'interazione AB

$$H_0 : \tau_i = 0, i = 1,\dots,a$$

$$H_0 : \beta_j = 0,\ j = 1,\dots,b$$

$$H_0 : (\tau\beta)_{ij} = 0 \begin{cases} i = 1,..,a \\ j = 1,\dots,b \end{cases}$$

$$y_{i..} = \sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{i..} = \frac{y_{i..}}{bn} \quad i = 1,2,\dots,a$$

$$y_{.j.} = \sum_{i=1}^{a}\sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{.j.} = \frac{y_{.j.}}{an} \quad j = 1,2,\dots,b$$

$$y_{ij.} = \sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{ij.} = \frac{y_{ij.}}{n} \quad \begin{matrix} i = 1,2,\dots,a \\ j = 1,2,\dots,b \end{matrix}$$

$$y_{...} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk} \quad \bar{y}_{...} = \frac{y_{...}}{abn}$$

		Fattore B				Media
		1	2	…	b	$y_{i..}$
Fattore A	1	$y_{111}, y_{112}, \dots y_{21n}$	$y_{121}, y_{122}. \dots y_{12n}$	…	$y_{1b1}. y_{1b2}, \dots y_{1bn}$	$y_{1..}$
	2	$y_{211}, y_{212}, \dots y_{21n}$	$y_{221}, y_{222}, \dots y_{22n}$	…	$y_{2b1}, y_{2b2}, \dots y_{2bn}$	$y_{2..}$
	a	$y_{a11}, y_{a12}, \dots y_{11n}$	$y_{a21}, y_{a22}, \dots y_{a2n}$	…	$y_{ab1}, y_{ab2}, \dots y_{abn}$	$y_{a..}$
	Media $y_{.j.}$	$y_{.1.}$	$y_{.2.}$		$y_{.b.}$	$y_{...}$

$$SS_{totale} = SS_A + SS_B + SS_{AB} + SS_{errore} \qquad SS_{totale} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2$$

$$\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{...})^2 = bn\sum_{i=1}^{a}(\bar{y}_{i..} - \bar{y}_{...})^2 + an\sum_{j=1}^{b}(\bar{y}_{.j.} - \bar{y}_{...})^2 +$$

$$+ n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2 + \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(y_{ijk} - \bar{y}_{ij.})^2$$

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Media dei quadrati	F_0
A	SS_A	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$F_0 = \frac{MS_A}{MS_{errore}}$
B	SS_B	$b-1$	$MS_B = \frac{SS_B}{b-1}$	$F_0 = \frac{MS_B}{MS_{errore}}$
Interazione	SS_{AB}	$(a-1)(b-1)$	$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)}$	$F_0 = \frac{MS_{AB}}{MS_{errore}}$
Errore	SS_{errore}	$ab(n-1)$	$MS_{errrore} = \frac{SS_{errore}}{ab(n-1)}$	
Totale	SS_{totale}	$abn-1$		

Esempio: Le vernici di base per aereo sono applicate alla superficie di alluminio con due metodi: a immersione o a spruzzo. Lo scopo della vernice di base è di migliorare l'adesione della vernice: su alcune parti la vernice può essere applicata con entrambi i metodi. Un ingegnere, interessato a studiare tre differenti tipi di vernice rispetto alle loro differenti qualità di adesione, ha eseguito un esperimento fattoriale per investigare l'effetto del tipo di vernice e del metodo di applicazione sulla adesione della vernice. Tre campioni vengono verniciati con ciascun tipo di vernice usando ciascun metodo di applicazione, poi una vernice di finitura viene applicata, e la forza di adesione viene misurata. Le 18 prove di questo esperimento sono state eseguite in ordine casuale. Lo scopo dell'esperimento è determinare quale combinazione di vernice e di metodo di applicazione produce la massima forza di adesione. Sarebbe desiderabile che almeno uno dei tre tipi di vernice di base producesse una alta adesione, *indipendentemente* dal metodo di applicazione, perché ciò aggiungerebbe flessibilità al processo di fabbricazione.

	Metodo di applicazione		
Tipo di vernice	A immersione	A spruzzo	$y_{i..}$
1	4.0, 4.5, 4.3 12.8	5.4, 4.9, 5.6 15.9	28.7
2	5.6, 4.9, 5.4 15.9	5.8, 6.1, 6.3 18.2	34.1
3	3.8, 3.7, 4.0 11.5	5.5, 5.0, 5.0 15.5	27.0
$y_{.j.}$	40.2	49.6	$89.8 = y_{...}$

totali delle celle

$$SS_{\text{Totale}} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n} y_{ijk}^2 - \frac{y_{...}^2}{abn}$$

$$= (4.0)^2 + (4.5)^2 + \cdots + (5.0)^2 - \frac{(89.8)^2}{18} = 10.72$$

$$SS_{\text{Vernice}} = \sum_{i=1}^{a} \frac{y_{i..}^2}{bn} - \frac{y_{...}^2}{abn}$$

$$= \frac{(28.7)^2 + (34.1)^2 + (27.0)^2}{6} - \frac{(89.8)^2}{18} = 4.58$$

$$SS_{\text{Metodo}} = \sum_{j=1}^{b} \frac{y_{.j.}^2}{an} - \frac{y_{...}^2}{abn}$$

$$= \frac{(40.2)^2 + (49.6)^2}{9} - \frac{(89.8)^2}{18} = 4.91$$

$$SS_{\text{interazione}} = \sum_{i=1}^{a}\sum_{j=1}^{b} \frac{y_{ij.}^2}{n} - \frac{y_{...}^2}{abn} - SS_{\text{Vernice}} - SS_{\text{metodo}}$$

$$= \frac{(12.8)^2 + (15.9)^2 + (11.5)^2 + (15.9)^2 + (18.2)^2 + (15.5)^2}{3}$$

$$- \frac{(89.8)^2}{18} - 4.58 - 4.91 = 0.24$$

$$SS_{\text{Errore}} = SS_{\text{Totale}} - SS_{\text{Vernice}} - SS_{\text{Metodo}} - SS_{\text{interazione}}$$

$$= 10.72 - 4.58 - 4.91 - 0.24 = 0.99$$

$$F_{0A} > F_{0.05,2,12} = 3{,}89 \qquad F_{0B} > F_{0.05,1,12} = 4{,}75$$

$$F_{0AB} < F_{0.05,2,12} = 3{,}89$$

Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Media dei quadrati	F_0	P
Tipo di vernice	4.58	2	2.29	28.63	2.71×10^{-5}
Metodo	4.91	1	4.91	61.38	4.65×10^{-5}
Interazione	0.24	2	0.12	1.5	0.27
Errore	0.99	12	0.08		
Totale	10.72	17			

Gli effetti di A e B sono significativi, mentre non è presente interazione tra i due fattori

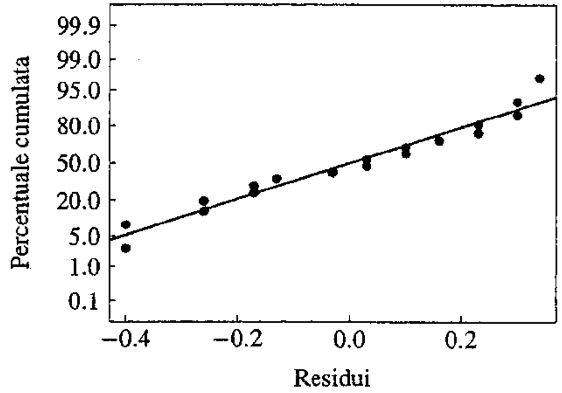
Analisi dei residui

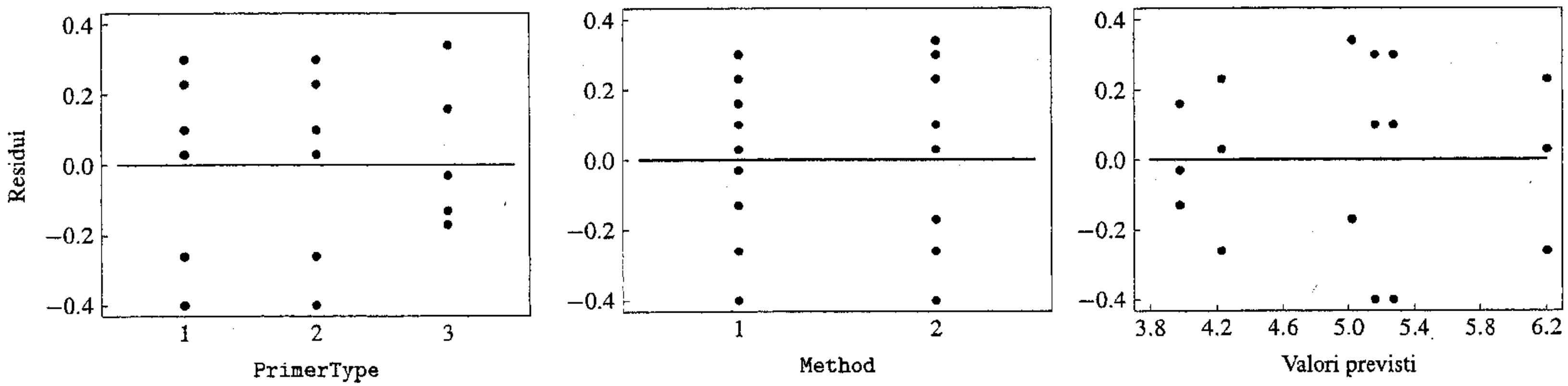
I residui svolgono un ruolo importante nella valutazione della adeguatezza del modello, verificando che gli errori del modello siano normalmente ed indipendentemente distribuiti, con la stessa varianza al variare del livello dei fattori.

$$e_{ijk} = y_{ijk} - \hat{y}_{ijk} =$$
$$= y_{ijk} - \bar{y}_{ij.}$$

→ assunzione di <u>normalità</u>

→ assunzione di <u>uguaglianza</u> delle <u>varianze</u>



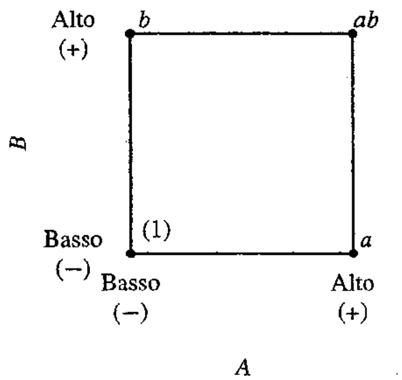


Piani fattoriali 2^k

Piani con k fattori ciascuno dei quali ha due livelli (indicati "-" e "+" e chiamati "basso" e "alto"). Dato che ogni replicazione completa del piano ha 2^k prove, tale disposizione si chiama piano fattoriale 2^k. Questi piani hanno un'analisi molto semplificata.

Il piano 2^2

Si definisce effetto principale del fattore la variazione nella risposta prodotta dal cambiamento di livello del fattore.
Si ha interazione tra fattori quando la differenza di risposta fra i livelli di un fattore non è la stessa a tutti i livelli degli altri fattori.



Effetto principale di A

$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{a+ab}{2n} - \frac{b+(1)}{2n} = \frac{1}{2n}\left[a + ab - b - (1)\right]$$

Effetto principale di B

$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{b+ab}{2n} - \frac{a+(1)}{2n} = \frac{1}{2n}\underbrace{\left[b + ab - a - (1)\right]}_{\text{Contrasto}}$$

Interazione AB

$$AB = \frac{ab+(1)}{2n} - \frac{a+b}{2n} = \frac{1}{2n}\left[ab + (1) - a - b\right]$$

Segni per gli effetti nel piano 2^2

Prova		Effetti fattoriali			
		I	A	B	AB
1	(1)	+	-	-	+
2	a	+	+	-	-
3	b	+	-	+	-
4	ab	+	+	+	+

Proprietà:

- Eccetto che per la colonna identità I ogni colonna ha un egual numero di segni più e meno;
- La somma dei prodotti dei segni di ogni coppia di colonne è zero; cioè le colonne sono ortogonali;
- Moltiplicando ogni colonna per la colonna I la si lascia invariata; cioè I è l'elemento identità;
- Il prodotto di ogni coppia di colonne produce una colonna della tabella (ad es. A x B = AB); ogni colonna moltiplicata per se stessa è la colonna identità.

La stima di ogni effetto o interazione potrà essere determinata così:

$$Effetto = \frac{Contrasto}{2n}$$

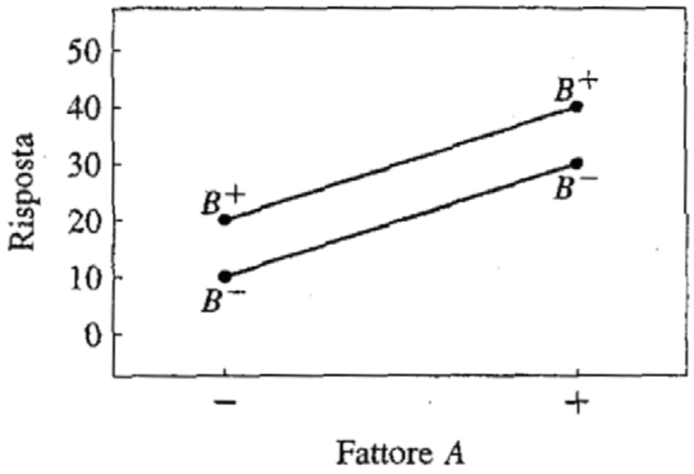
(per un piano 2^2)

La somma dei quadrati di ogni effetto è:

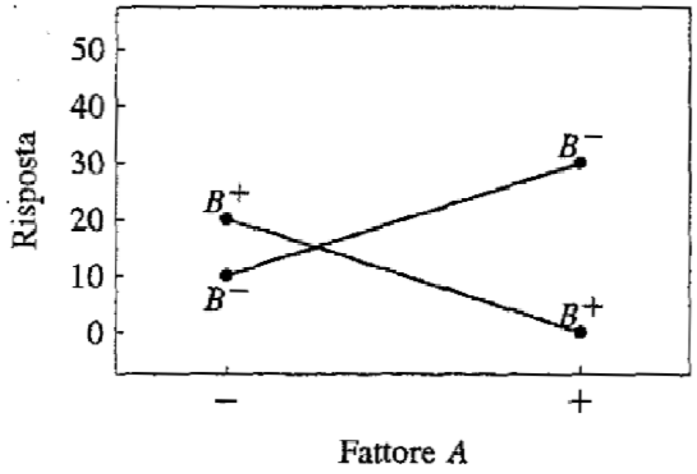
$$SS = \frac{(Contrasto)^2}{4n}$$

(per un piano 2^2)

Esperimento fattoriale <u>senza</u> <u>interazione</u>:
le rette sono approssimativamente parallele.

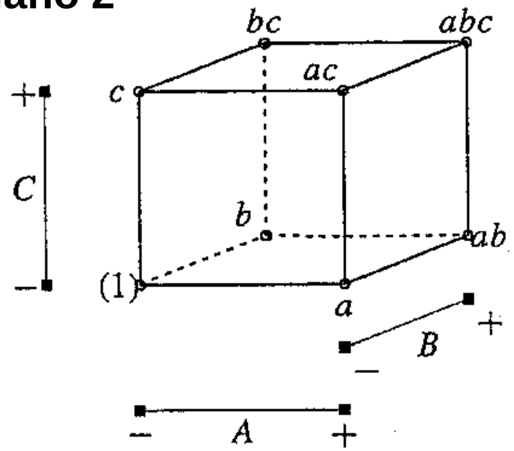


Esperimento fattoriale <u>con</u> <u>interazione</u>:
le rette non sono parallele.



Interazione fra i fattori in un piano fattoriale

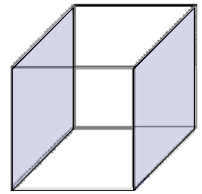
Il piano 2^3



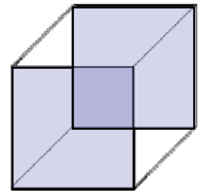
Modello fattoriale completo:

$$y = \underbrace{\mu}_{\text{media generale}} + \underbrace{A + B + C}_{\text{effetti principali}} + \underbrace{AB + AC + BC + ABC}_{\text{effetti di interazione tra fattori}} + \underbrace{\varepsilon}_{\text{errore casuale}}$$

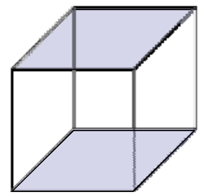
$$A = \bar{y}_{A^+} - \bar{y}_{A^-} = \frac{1}{4n}\left[a + ab + ac + abc - b - c - bc - (1)\right]$$



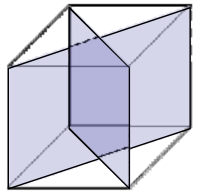
$$B = \bar{y}_{B^+} - \bar{y}_{B^-} = \frac{1}{4n}\left[b + ab + bc + abc - a - c - ac - (1)\right]$$



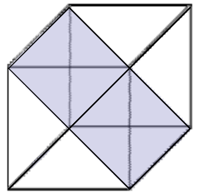
$$C = \bar{y}_{C^+} - \bar{y}_{C^-} = \frac{1}{4n}\left[c + ac + bc + abc - a - b - ab - (1)\right]$$



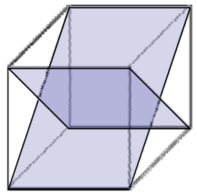
$$AB = \frac{1}{4n}\left[ab + (1) + abc + c - b - a - bc - ac\right]$$



$$AC = \frac{1}{4n}\left[ac + (1) + abc + b - c - a - ab - bc\right]$$

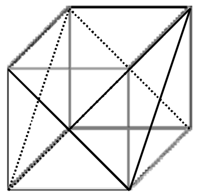


$$BC = \frac{1}{4n}\left[bc + (1) + abc + a - b - c - ab - ac\right]$$



$$ABC = \frac{1}{4n}\left[\left[abc - bc\right] - \left[ac - c\right] - \left[ab - b\right] + \left[a - (1)\right]\right] =$$

$$= \frac{1}{4n}\left[abc - bc - ac + c - ab + b + a - (1)\right]$$



NB: ABC è la stima dell'effetto dell'interazione fra i tre fattori (A,B,C) del piano fattoriale 2^3.

Prova		Effetti fattoriali							
		I	A	B	AB	C	AC	BC	ABC
1	(1)	+	-	-	+	-	+	+	-
2	a	+	+	-	-	-	-	+	+
3	b	+	-	+	-	-	+	-	+
4	ab	+	+	+	+	-	-	-	-
5	c	+	-	-	+	+	-	-	+
6	ac	+	+	-	-	+	+	-	-
7	bc	+	-	+	-	+	-	+	-
8	abc	+	+	+	+	+	+	+	+

Segni per gli effetti nel piano 2^3

Relazioni valide per un piano fattoriale 2^k

La stima di ogni effetto o interazione potrà essere determinata così:

$$Effetto = \frac{Contrasto}{n2^{k-1}}$$

La somma dei quadrati di ogni effetto è:

$$SS = \frac{(Contrasto)^2}{n2^k}$$

Procedura di analisi per gli esperimenti fattoriali

Stimare gli effetti dei fattori

↓

Costruire il modello iniziale → in ogni esperimento fattoriale con replicazioni si usa quasi sempre il modello fattoriale completo

↓

Testare la significatività degli effetti → mediante Analisi della Varianza

↓

Analizzare i residui

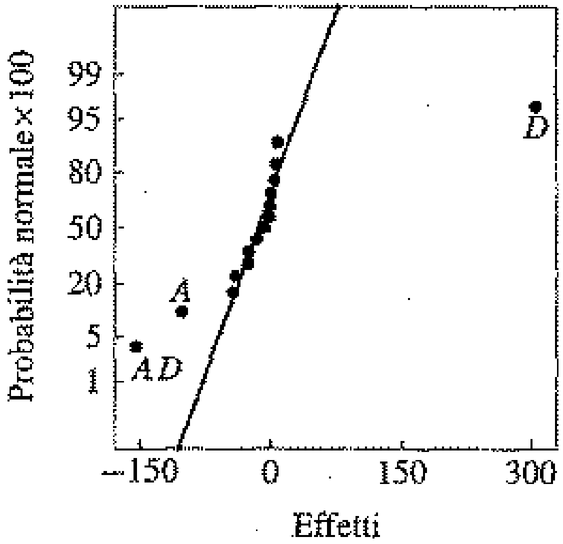
↓

Migliorare il modello, se necessario → ad esempio si raffina il modello mediante eliminazione di termini non significativi dal modello finale

↓

Interpretare i risultati

Un metodo molto utile per valutare la significatività dei fattori in un esperimento 2^k è la costruzione di un grafico probabilistico delle stime degli effetti.



Se nessuno degli effetti è significativo, le stime si comportano come un campione casuale estratto da una distribuzione normale con media nulla; inoltre gli effetti riportati sul grafico giaceranno approssimativamente su di una linea retta.

Quegli effetti che si discostano dalla linea retta sono fattori significativi.

Modelli di regressione e analisi dei residui

Modello di regressione Valido aiuto nell'interpretazione dell'esperimento e per prevedere la risposta di un sistema.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots\dots\dots + \beta_k x_k + \varepsilon$$

dove *y* è la variabile di risposta, le *x* sono un'insieme di regressori o variabili predittive, le β sono i coefficienti di regressione, ε è un termine di errore casuale, assunto indipendente e di legge normale con media nulla e varianza σ^2.

Nel caso particolare di un piano a 2 fattori, la cui interazione è significativa, il modello di regressione da adottare risulta:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

Nel caso speciale dei piani 2^k, è estremamente semplice determinare le stime ai minimi quadrati delle β: la stima ai minimi quadrati di ogni coefficiente di regressione β è semplicemente la metà della corrispondente stima dell'effetto del fattore.

Si faccia attenzione che il risultato precedente è basato sull'assunto che le variabili *x* che rappresentano i fattori del piano siano codificate con gli estremi dell'intervallo $-1 \le x \le 1$.

Prima di poter trarre conclusioni sul modello adottato occorre verificare l'adeguatezza nel spiegare le variazioni nella risposta. Lo strumento che permette di fare questa verifica è l'analisi dei residui.

I residui vengono calcolati come differenze tra le osservazioni effettivamente ottenute ed i valori previsti (calcolati attraverso il modello di regressione ai dati).

Residuo: $e_{ijk} = y_{ijk} - \hat{y}_{ijk} =$

$$= y_{ijk} - \bar{y}_{ij.}$$

Grafico dei residui in funzione del valore stimato dal modello

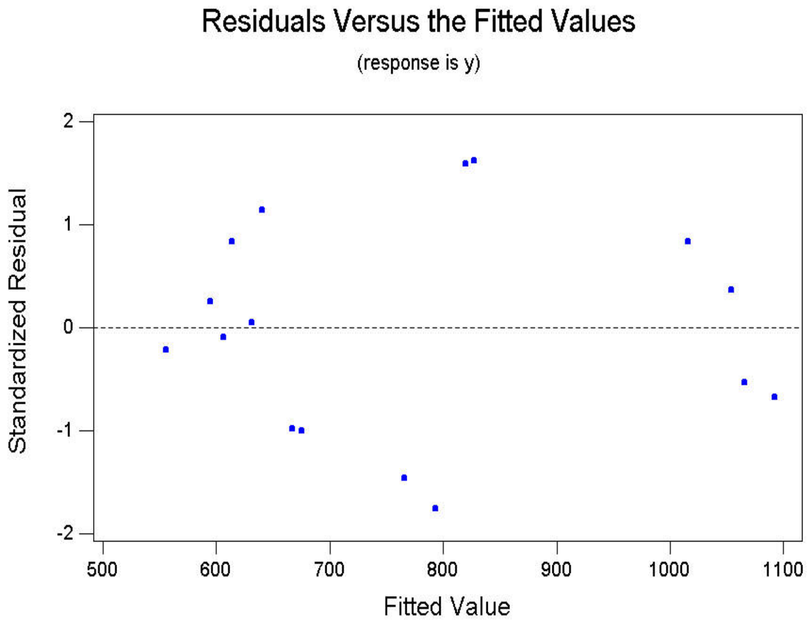
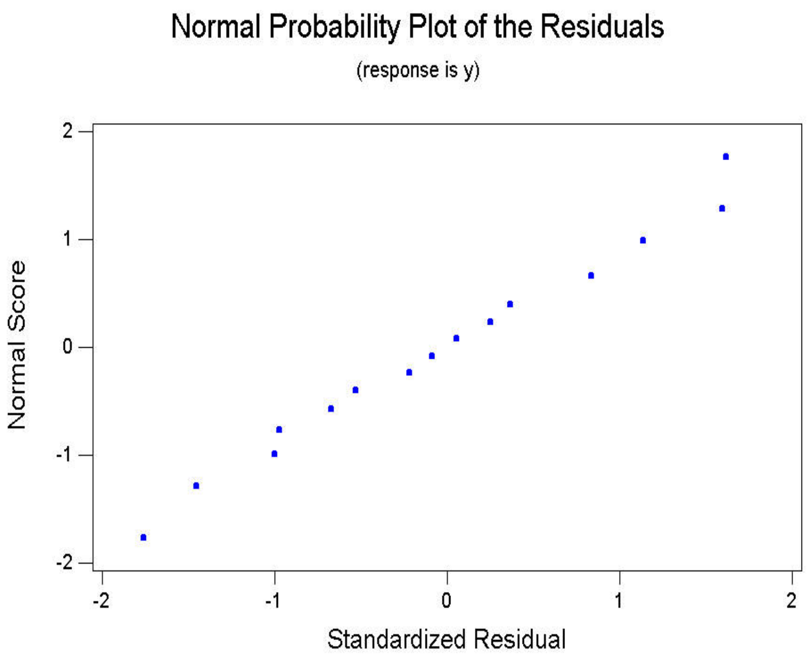


Grafico in scala normale dei residui



Se i residui cadono approssimativamente su una retta passante per il centro del grafico, allora verifico l'assunzione di normalità degli errori nel modello.

Esempio:

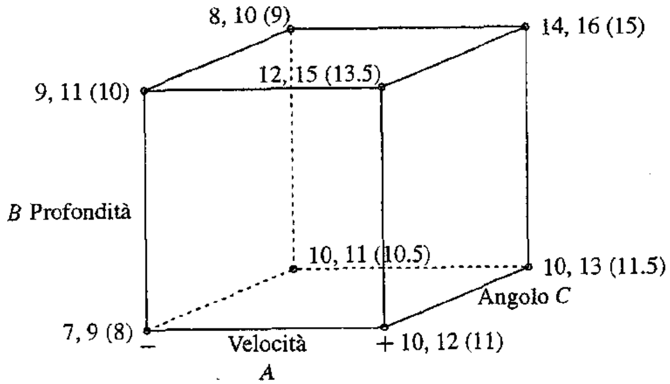
È stato eseguito un esperimento per studiare la finitura superficiale di un pezzo metallico. L'esperimento è un piano fattoriale 2^3 nei fattori velocità di avanzamento (A), profondità di passata (B) e angolo dell'attrezzo (C), con *n = 2* replicazioni.

$$A = \frac{1}{4n}[a + ab + ac + abc - b - c - bc - (1)]$$
$$= \frac{1}{4(2)}[22 + 27 + 23 + 30 - 20 - 21 - 18 - 16]$$
$$= \frac{1}{8}[27] = 3.375$$

$$SS_A = \frac{(\text{Contrasto}_A)^2}{n2^k} = \frac{(27)^2}{2(8)} = 45.5625$$

$B = 1.625$	$SS_B = 10.5625$
$C = 0.875$	$SS_C = 3.0625$
$AB = 1.375$	$SS_{AB} = 7.5625$
$AC = 0.125$	$SS_{AC} = 0.0625$
$BC = -0.625$	$SS_{BC} = 1.5625$
$ABC = 1.125$	$SS_{ABC} = 5.5625$

Prova		Fattori del piano A	B	C	Finitura della superficie	Totali
1	(1)	−1	−1	−1	9, 7	16
2	*a*	1	−1	−1	10, 12	22
3	*b*	−1	1	−1	9, 11	20
4	*ab*	1	1	−1	12, 15	27
5	*c*	−1	−1	1	11, 10	21
6	*ac*	1	−1	1	10, 13	23
7	*bc*	−1	1	1	10, 8	18
8	*abc*	1	1	1	16, 14	30



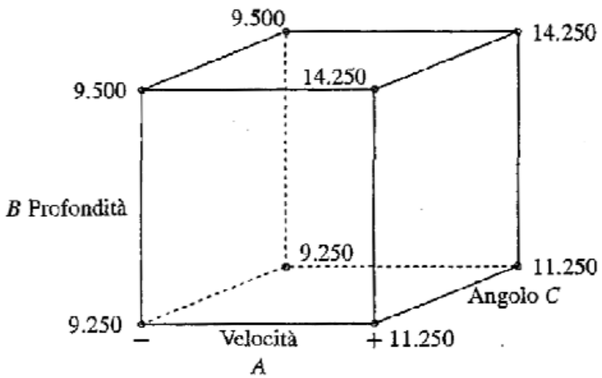
Sorgente di variazione	Somma dei quadrati	Gradi di libertà	Media dei quadrati	F_0	P
A	45.5625	1	45.5625	18.69	2.54×10^{-3}
B	10.5625	1	10.5625	4.33	0.07
C	3.0625	1	3.0625	1.26	0.29
AB	7.5625	1	7.5625	3.10	0.12
AC	0.0625	1	0.0625	0.03	0.88
BC	1.5625	1	1.5625	0.64	0.45
ABC	5.5625	1	5.5625	2.08	0.19
Errore	19.5000	8	2.4375		
Totale	92.9375	15			

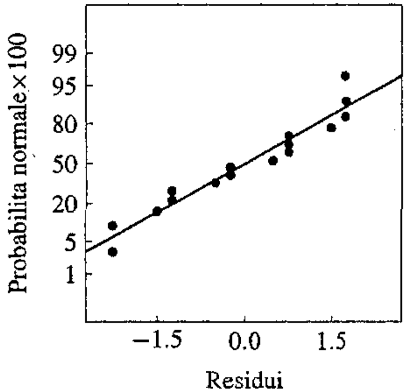
Considero significativi l'effetto A, l'effetto B ed infine la loro interazione AB

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$$

$$\hat{y} = 11.0625 + \left(\frac{3.375}{2}\right) x_1 + \left(\frac{1.625}{2}\right) x_2 + \left(\frac{1.375}{2}\right) x_1 x_2$$
$$= 11.0625 + 1.6875 x_1 + 0.8125 x_2 + 0.6875 x_1 x_2$$

NB: VALE SOLO PER PIANI FATTORIALI A DUE LIVELLI





Rappresentazione grafica di un modello di regressione

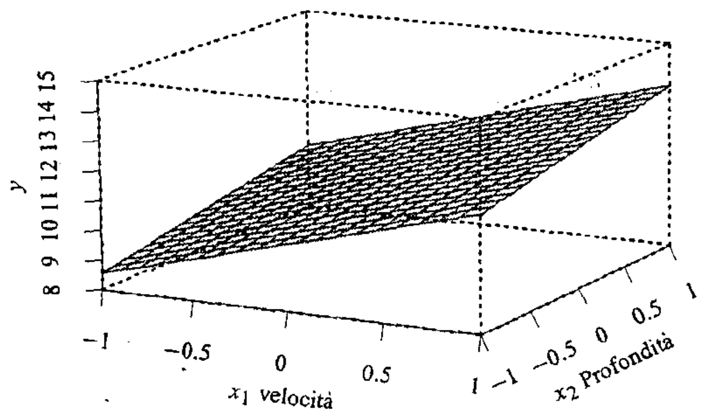
Si graficano i valori della risposta y previsti dal modello di regressione ai dati, al fine di ottenere un valido aiuto nella interpretazione dei risultati dell'esperimento.

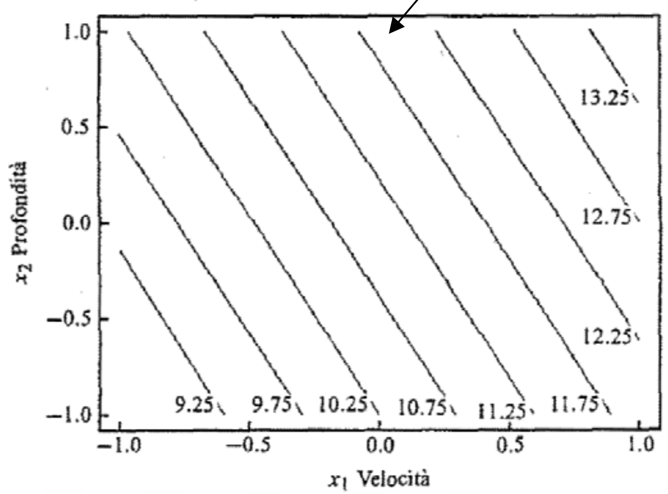
La rappresentazione grafica dei valori delle risposte previste prende il nome di grafico della superficie di risposta ed il modello di regressione utilizzato per generare questo grafico è spesso chiamato modello di superficie di risposta.

In presenza di linearità negli effetti dei fattori, si utilizza un modello di superficie di risposta del primo ordine.

Nel caso di *k* = *2* fattori: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$

curva di livello



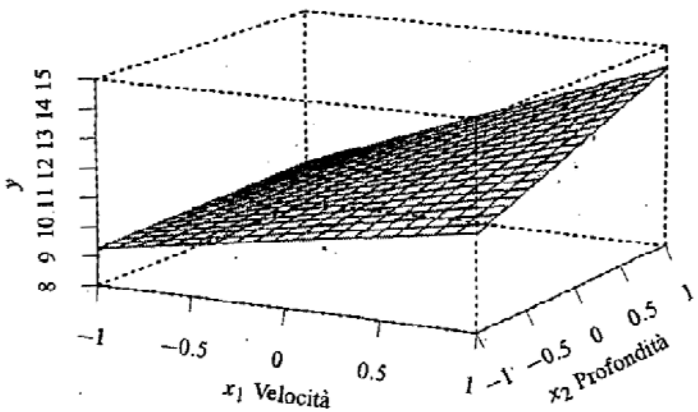


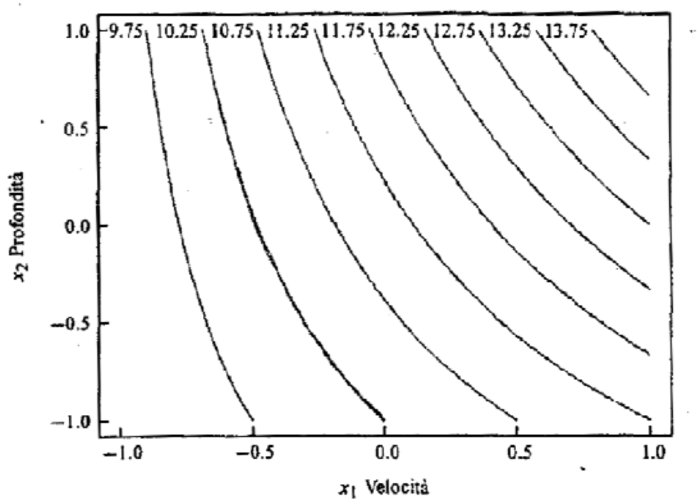
Il grafico bidimensionale delle curve di livello si ottiene guardando dall'alto il grafico tridimensionale della superficie di risposta e connettendo i punti nel piano x_1-x_2 che hanno lo stesso valore di risposta prevista y.

In presenza di effetti di interazione tra i diversi fattori, si utilizza un modello di superficie di risposta del primo ordine con interazioni.

Nel caso di *k = 2* fattori: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon$

L'effetto della aggiunta del termine di interazione è quello di introdurre una curvatura nella superficie di risposta: in effetti il piano è "torto" dall'effetto di interazione.
Inoltre le linee di risposta costante (curve di livello) non sono più linee rette bensì curve.





Il piano 2^k senza replicazioni

Con la crescita del numero di fattori in un esperimento fattoriale, cresce anche il numero di effetti che possono essere stimati: ad esempio in un esperimento 2^6 si hanno 6 effetti semplici, 15 interazioni fra due fattori, 20 interazioni fra tre fattori, 15 interazioni fra quattro fattori, 6 interazioni fra cinque fattori ed 1 interazione fra sei fattori.

In molte applicazioni si applica allora il principio di economia *sparsity* degli effetti: il sistema è cioè controllato dagli effetti di principali e dalle interazioni di ordine basso; le interazioni fra tre o più fattori sono in genere trascurabili.

Quando il numero di fattori è dunque molto grande (k ≥ 4 o 5), è pratica comune di eseguire il piano 2^k solo una volta, cioè senza replicazioni, e poi raggruppare o combinare le stime delle interazioni di ordine alto come stima dell'errore.

$$SS_{Errore} = \sum SS_{\substack{\text{Interazioni di} \\ 3\,\text{o} + \text{fattori}}}$$

$$g.d.l._{Errore} = \sum g.d.l._{\substack{\text{Interazioni di} \\ 3\,\text{o} + \text{fattori}}} \longrightarrow MS_{Errore} = \frac{SS_{Errore}}{g.d.l._{Errore}}$$

Aggiunta dei punti centrali ad un piano 2^k

Una potenziale fonte di preoccupazione nell'uso di piani fattoriali a due livelli è l'ipotesi di linearità negli effetti dei fattori.

Una linearità perfetta non è necessaria, e i piani 2^k funzioneranno piuttosto bene quando l'ipotesi di linearità è verificata solo approssimativamente: è possibile infatti aggiungere un termine di interazione al modello di soli effetti semplici, ottenendo così una certa protezione contro la curvatura .

In alcuni sistemi o processi è necessario includere **effetti del secondo ordine** al fine di ottenere un modello adeguato.

Nel caso di $k = 2$ fattori:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \underbrace{\beta_{11} x_1^2 + \beta_{22} x_2^2}_{\text{effetti quadratici puri}} + \varepsilon$$

Per adottare il modello quadratico, è necessario che tutti i fattori compaiano nelle prove con almeno tre livelli.

Modello quadratico per k fattori:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{i<j}\sum \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \varepsilon$$

Per verificare se i termini quadratici puri sono necessari, si utilizza un metodo che consiste nell'aggiunta di punti centrali al piano 2^k.

I punti centrali sono costituiti da n_C replicazioni nel punto $x_i = 0$ (i = 1, 2, …, k).

L'aggiunta di tali punti al centro non interferisce con le usuali stime degli effetti nel piano 2^k.

Test d'ipotesi per la curvatura:

$$H_0 : \sum_{j=1}^{k} \beta_{jj} = 0$$

$$H_1 : \sum_{j=1}^{k} \beta_{jj} \neq 0$$

$$F_0 = \frac{MS_{Quadratica}}{MS_{Errore}}$$

Somma di quadrati con un grado di libertà per la curvatura quadratica pura

$$SS_{Quadratica} = \frac{n_F n_C (\bar{y}_F - \bar{y}_C)^2}{n_F + n_C}$$

$\bar{y}_C \longrightarrow$ media delle n_C osservazioni nel punto centrale

$\bar{y}_F \longrightarrow$ media delle prove negli n_F punti fattoriali

$n_F \longrightarrow$ numero di punti nel piano fattoriale

Inoltre, se i punti del piano fattoriale non sono replicati, è possibile utilizzare gli n_C punti centrali per costruire una stima dell'errore con n_C -1 gradi di libertà.