

Analisi della varianza (anova) a due vie

Andrea Onofri

27 marzo 2014

Indice

1	Il concetto di 'interazione'	1
2	Tipi di interazione	2
3	Descrizione del caso studio	3
4	Analisi dei dati	4

Sommario

Scopo di questo documento è illustrare come vengono pianificati, organizzati ed analizzati gli esperimenti fattoriali, con due tesi sperimentali a confronto.

1 Il concetto di 'interazione'

In molti casi il ricercatore è interessato ad organizzare esperimenti per valutare l'effetto di due fattori sperimentali combinati (ad esempio la lavorazione del terreno ed il diserbo chimico), in modo da mettere in evidenza possibili "interazioni". Con questo termine si vuole intendere il fatto che l'effetto di un fattore (ad es. la lavorazione) cambia a seconda dei livelli dell'altro fattore (il diserbo chimico) come ad esempio nella tabella 1.

Nella tabella sovrastante osserviamo che il trattamento A1 ha incrementato il risultato di 1.75 unità rispetto alla media generale, mentre il trattamento B1 ha incrementato il risultato di 0.25 unità rispetto alla media generale. Di conseguenza, ci aspetteremmo per la combinazione A1B1 un risultato finale pari a $11.75 + 1.75 + 0.25 = 13.75$, mentre il risultato finale è di 10 unità. Evidentemente la combinazione A1B1 è una combinazione svantaggiosa, cioè i due trattamenti interagiscono tra di loro in modo negativo ed il risultato è inferiore alle attese.

Tabella 1: Esempio di interazione tra fattori sperimentali: le medie di B1 e B2 (medie marginali) sono praticamente simili, ma ciò è dovuto al fatto che gli effetti cambiano al cambiare del livello di A (A1 e A2), come si può vedere dalle medie di cella

Fattori sperimentali	B1	B2	Media	Effetto
A1	10	17	13.5	1.75
A2	14	6	10	-1.75
Media	12	11.5	11.75	
Effetto	0.25	-0.25	11.75	

2 Tipi di interazione

In genere, abbiamo due tipi di interazione: quella in cui cambia la graduatoria tra i trattamenti (CROSSOVER INTERACTION, Fig. 2 e tab. 1) o quella in cui vi è solo una modifica dell'entità dell'effetto (NON-CROSSOVER INTERACTION; Fig. 2). Questi aspetti possono evidenziarsi solamente con esperimenti combinati a più fattori (esperimenti fattoriali) e mai con esperimenti singoli.

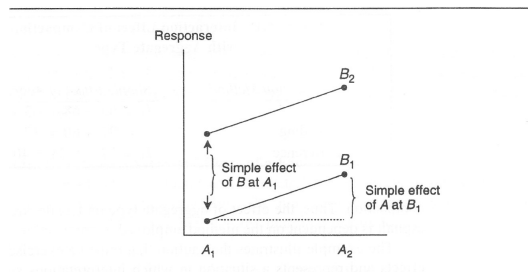


Figure 6.1 Illustration of no interaction in a factorial arrangement

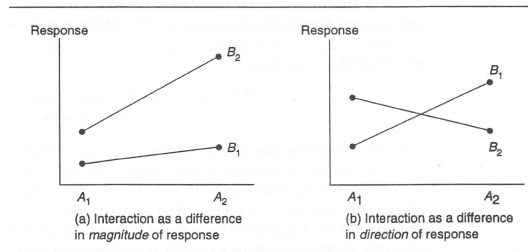


Figure 6.2 Illustration of interaction in a factorial arrangement

Figura 1: Esempi di interazione tra fattori sperimentali (Da Kuhel, 2000)

3 Descrizione del caso studio

Un ricercatore ha organizzato un esperimento fattoriale dove ha valutato l'effetto di tre tipi di lavorazione del terreno (lavorazione minima=LM aratura superficiale=SUP ed aratura profonda=PROF) e di due tipi di diserbo chimico (a pieno campo=TOT e localizzato sulla fila della coltura=PARZ). L'ipotesi scientifica è che, in caso di diserbo localizzato, il rovesciamento del terreno prodotto dall'aratura sia fondamentale in quanto sotterra i semi prodotti dalle piante infestanti, impedendone l'emergenza nella coltura e massimizzando le rese.

<i>MIN</i> <i>PARZ</i>	<i>SUP</i> <i>PARZ</i>	<i>MIN</i> <i>PARZ</i>	<i>PROF</i> <i>TOT</i>
<i>PROF</i> <i>TOT</i>	<i>PROF</i> <i>PARZ</i>	<i>SUP</i> <i>TOT</i>	<i>PROF</i> <i>PARZ</i>
<i>SUP</i> <i>PARZ</i>	<i>MIN</i> <i>PARZ</i>	<i>PROF</i> <i>TOT</i>	<i>SUP</i> <i>TOT</i>
<i>PROF</i> <i>PARZ</i>	<i>SUP</i> <i>TOT</i>	<i>SUP</i> <i>PARZ</i>	<i>MIN</i> <i>TOT</i>
<i>SUP</i> <i>TOT</i>	<i>MIN</i> <i>TOT</i>	<i>PROF</i> <i>PARZ</i>	<i>SUP</i> <i>PARZ</i>
<i>MIN</i> <i>TOT</i>	<i>PROF</i> <i>TOT</i>	<i>MIN</i> <i>TOT</i>	<i>MIN</i> <i>PARZ</i>

Figura 2: Esempio di un disegno sperimentale fattoriale a blocchi randomizzati. I colori contraddistinguono i quattro blocchi.

In totale, l'esperimento include sei tesi (le sei possibili combinazioni tra i due fattori sperimentali) e quattro repliche per un totale di 24 parcelle. Come consuetudine in pieno campo, l'esperimento è organizzato a blocchi randomizzati e le sei tesi sperimentali sono randomizzate all'interno di ciascun blocco (figura 2).

I risultati ottenuti con questo esperimento sono riportati nella tabella 2.

Tabella 2: Risultati ottenuti in un esperimento fattoriale

Lavorazione	Diserbo	Blocco	Produzione zucchero (t/ha)
MIN	tot	1	11.614
MIN	tot	2	9.283
MIN	tot	3	7.019
MIN	tot	4	8.015
MIN	parz	1	5.117
MIN	parz	2	4.306
MIN	parz	3	8.940
MIN	parz	4	5.617
SUP	tot	1	10.019
SUP	tot	2	8.686
SUP	tot	3	8.904
SUP	tot	4	8.956
SUP	parz	1	7.975
SUP	parz	2	8.355
SUP	parz	3	7.990
SUP	parz	4	9.581
PROF	parz	1	11.477
PROF	parz	2	10.671
PROF	parz	3	10.567
PROF	parz	4	9.800
PROF	tot	1	10.111
PROF	tot	2	8.784
PROF	tot	3	7.927
PROF	tot	4	10.005

4 Analisi dei dati

Il primo passo dell'analisi è il calcolo delle medie, che porta a risultati accettabili, dato che il disegno è perfettamente bilanciato (ugual numero di repliche per ogni combinazione dei fattori sperimentali). La media generale è pari a 8.738 (24 valori), mentre le medie per la lavorazioni (8 valori ognuna) sono:

$$\begin{aligned} \text{MIN} &= 7.489 \\ \text{SUP} &= 8.808 \\ \text{PROF} &= 9.918 \end{aligned}$$

Le medie del diserbo (12 valori ognuna) sono:

$$\begin{aligned} \text{TOT} &= 9.110 \\ \text{PARZ} &= 8.366 \end{aligned}$$

Oltre a ciò, è possibile calcolare la media per ogni combinazione lavorazione x diserbo (quattro valori ognuna):

$$\text{MIN} - \text{TOT} = 8.983$$

$$\text{MIN} - \text{PARZ} = 5.995$$

$$\text{SUP} - \text{TOT} = 9.141$$

$$\text{SUP} - \text{PARZ} = 8.475$$

$$\text{PROF} - \text{TOT} = 10.629$$

$$\text{PROF} - \text{PARZ} = 9.207$$

Oltre alle medie anzidette abbiamo anche le medie dei blocchi (sei valori ognuna), che sono:

$$1 - 9.39$$

$$2 - 8.35$$

$$3 - 8.56$$

$$4 - 8.66$$

Nella analisi della varianza, abbiamo gli effetti principali (blocco, lavorazione e diserbo), più l'interazione, che misura se e quanto l'effetto di un fattore dipende dai livelli dell'altro.

Gli effetti principali (blocco, lavorazione e diserbo) sono facili da calcolare, in quanto risultano come differenza tra le medie e la media generale. L'effetto dell'interazione risulta invece dalla differenza tra la media delle combinazioni e la media generale addizionata degli effetti principali:

$$\text{Effetto interazione} = \text{Media combinazione} - (\text{Media generale} + \text{Effetto Lavorazioni} + \text{Effetto Diserbo})$$

Ad esempio, per la combinazione MIN - TOT, l'effetto è pari a:

$$8.983 - (8.738 - 1.249 + 0.372) = 1.122$$

La media generale addizionata degli effetti principali (i valori tra parentesi nell'equazione sovrastante) rappresenta la produzione attesa se gli effetti della lavorazione e del diserbo fossero completamente additivi (assenza di interazione).

La tabella degli effetti risulta di seguito.

Produzione osservata	Media generale	Effetto Lavorazione	Effetto Diserbo	Effetto Lav. x Dis.	Effetto Blocco	Residuo
11.614	8.738	-1.249	0.372	1.122	0.647	1.984
9.283	8.738	-1.249	0.372	1.122	-0.391	0.691
7.019	8.738	-1.249	0.372	1.122	-0.18	-1.784
8.015	8.738	-1.249	0.372	1.122	-0.076	-0.892
5.117	8.738	-1.249	-0.372	-1.122	0.647	-1.526
4.306	8.738	-1.249	-0.372	-1.122	-0.391	-1.298
8.94	8.738	-1.249	-0.372	-1.122	-0.18	3.126
5.617	8.738	-1.249	-0.372	-1.122	-0.076	-0.302
10.019	8.738	0.07	0.372	-0.039	0.647	0.23
8.686	8.738	0.07	0.372	-0.039	-0.391	-0.065
8.904	8.738	0.07	0.372	-0.039	-0.18	-0.057
8.956	8.738	0.07	0.372	-0.039	-0.076	-0.109
7.975	8.738	0.07	-0.372	0.039	0.647	-1.147
8.355	8.738	0.07	-0.372	0.039	-0.391	0.27
7.99	8.738	0.07	-0.372	0.039	-0.18	-0.305
9.581	8.738	0.07	-0.372	0.039	-0.076	1.182
11.477	8.738	1.179	-0.372	1.083	0.647	0.201
10.671	8.738	1.179	-0.372	1.083	-0.391	0.433
10.567	8.738	1.179	-0.372	1.083	-0.18	0.119
9.8	8.738	1.179	-0.372	1.083	-0.076	-0.753
10.111	8.738	1.179	0.372	-1.083	0.647	0.257
8.784	8.738	1.179	0.372	-1.083	-0.391	-0.032
7.927	8.738	1.179	0.372	-1.083	-0.18	-1.1
10.005	8.738	1.179	0.372	-1.083	-0.076	0.874

Per ognuno degli effetti inseriti nel modello è possibile calcolare la relativa devianza, giungendo alla tabella ANOVA seguente:

ANOVA

FATTORE	DEVIANZA	DF	VARIANZA	F	ProbF
Blocco	3.66	3	1.22	0.65	0.59
Lavorazione	23.65	2	11.82	6.32	0.01
Diserbo	3.32	1	3.32	1.78	0.20
Lavorazione x Diserbo	19.46	2	9.73	5.20	0.02
Residuo	28.06	15	1.87		
Totale	78.17	23	3.40		

Si può notare come lavorazione e diserbo interagiscono significativamente tra loro ($p < 0.05$).

DA RICORDARE SEMPRE

Nel leggere una tabella ANOVA a due (o più) vie, è **fondamentale procedere dal basso verso l'alto**, in quanto la presenza di un'interazione significativa rende non-informative sia le significanze degli effetti principali, sia le medie marginali. Infatti, come abbiamo visto, vi possono essere casi in cui le medie marginali sono simili, ma ciò è dovuto alla presenza di una interazione CROSSOVER.

Diagnostica

- Per controllare gli assunti di base con DSAASTAT, immaginiamo che il disegno sia a blocco randomizzato.
- Creiamo una nuova variabile, che combini i livelli delle due variabili precedenti (somma a stringa)
- ESEMPIO: nella cella A1 ho l'etichetta MIN e nella cella A2 ho PARZ. Immetto in A3 la formula = A1&A2, che mi restituisce MINPARZ
- Utilizzando questa nuova variabile, notiamo la presenza di un outlier (il settimo dato).
- DSAASTAT ci suggerisce la ricostruzione del dato adottando un valore più in linea con gli altri.
- Ignoriamo questo aspetto. Che cosa succederebbe se seguissimo il suggerimento?

Errori standard

- Abbiamo un errore standard per ogni effetto
- NUMERO DI REPLICHE: l'interazione ha un numero di repliche pari a quello effettivo, A ha un numero di repliche pari a quello effettivo per il numero dei livelli di B, B ha un numero di repliche pari a quello effettivo per il numero dei livelli di A

$$SEM_A = \sqrt{\frac{1.87}{4 \cdot 2}} = 0.483$$

$$SEM_B = \sqrt{\frac{1.87}{4 \cdot 3}} = 0.395$$

$$SEM_{AB} = \sqrt{\frac{1.87}{4}} = 0.684$$

Notare che le medie marginali, grazie al numero di repliche più elevato, sono stimate con maggiore precisione.

Dati sbilanciati

Valgono le considerazioni già fatte per il blocco randomizzato. Anche in questo caso, in presenza di pochi dati mancanti/aberranti si può procedere alla loro ricostruzione, con il metodo indicato per il blocco randomizzato. In caso

di molti dati mancanti e/o di intere combinazioni mancanti, è necessario utilizzare metodiche di analisi più avanzate, che non sono oggetto del presente corso.

MCP

- Le SED si ottengono dalle SEM, moltiplicandole per $\sqrt{2}$
- In questo caso, posso fare i confronti multipli solo tra le medie di cella (l'interazione è significativa)
- Ho solo sei medie, quindi posso utilizzare la MDS