



UNIVERSITÀ DEGLI STUDI DI MILANO

DIPARTIMENTO DI SCIENZE AGRARIE E AMBIENTALI
PRODUZIONE, TERRITORIO, AGROENERGIA

Metodologia Sperimentale Agronomica / Metodi Statistici per la Ricerca Ambientale

Marco Acutis

marco.acutis@unimi.it

www.acutis.it

a.a. 2018 - 2019

CdS Scienze della Produzione e Protezione delle Piante (g59)

CdS Biotecnologie Vegetali, Alimentari e Agro-Ambientali (g61)

CdS Scienze Agro-Ambientali (g57)

Lezione 03 - Sommario

- Analisi della Varianza
 - ANOVA a 2 o più vie



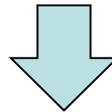
Analisi della Varianza - Il caso a 2 o più vie

Definizione

L'analisi della varianza a 1 via è lo schema più semplice di confronto simultaneo tra più medie. Nella pratica sperimentale spesso però rappresenta un'impostazione troppo elementare: infatti, in modo implicito, assume che tutta la variabilità presente nei diversi gruppi a confronto sia determinata dai differenti livelli del singolo fattore in osservazione.

Sovente è utile, quando non necessario, prendere in considerazione almeno due fonti di variabilità, allo scopo di:

- analizzare gli effetti di due o più cause contemporaneamente,
- ridurre la varianza d'errore, isolando gli effetti dovuti ad altre cause note.



Analisi della Varianza a 2 o più vie (o Fattoriale)

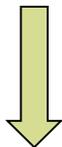


Analisi della Varianza - Il caso a 2 o più vie

Pro & Contro

Pro

- studio contemporaneo dell'effetto di 2 o più fattori
- identificazione dell'interazione tra fattori
- maggiore potenza



Contro

- difficoltà con esperimenti sbilanciati
- difficoltà d'interpretazione (se i fattori sono più di 2)
- complessità dei calcoli (trascurabile)



L'ANOVA a 2 o più vie è da preferire per esperimenti finalizzati all'identificazione dell'effetto di specifici trattamenti, mentre è meno efficace nell'analisi di sistemi reali.

Analisi della Varianza - Il caso a 2 o più vie

Interazione tra Fattori

Tra 2 (o più) fattori applicati contemporaneamente può esservi:

indifferenza



I fattori esercitano il loro effetto senza variazioni dovute al livello degli altri fattori (comportamento additivo).

sinergismo



La presenza contemporanea di determinati livelli dei fattori migliora il risultato rispetto alla semplice additività.

antagonismo



La presenza contemporanea di determinati livelli dei fattori peggiora il risultato rispetto alla semplice additività.

Comportamenti sinergici o antagonistici indicano **INTERAZIONE** tra i fattori.

Analisi della Varianza - Il caso a 2 o più vie

Esempio di Interazione (1/2)

Esempio (didattico) con 2 fattori, ciascun con 2 livelli

	Livello 0	Livello 1
N (Kg ha ⁻¹)	0	100
P ₂ O ₅ (Kg ha ⁻¹)	0	60

Le combinazioni possibili tra i livelli dei fattori sono N0-P0, N0-P1, N1-P0 e N1-P1.

Per ciascuna combinazione sono state effettuate 3 ripetizioni.

Le combinazioni vanno considerate come se fossero singoli trattamenti nell'ANOVA a 1 via.

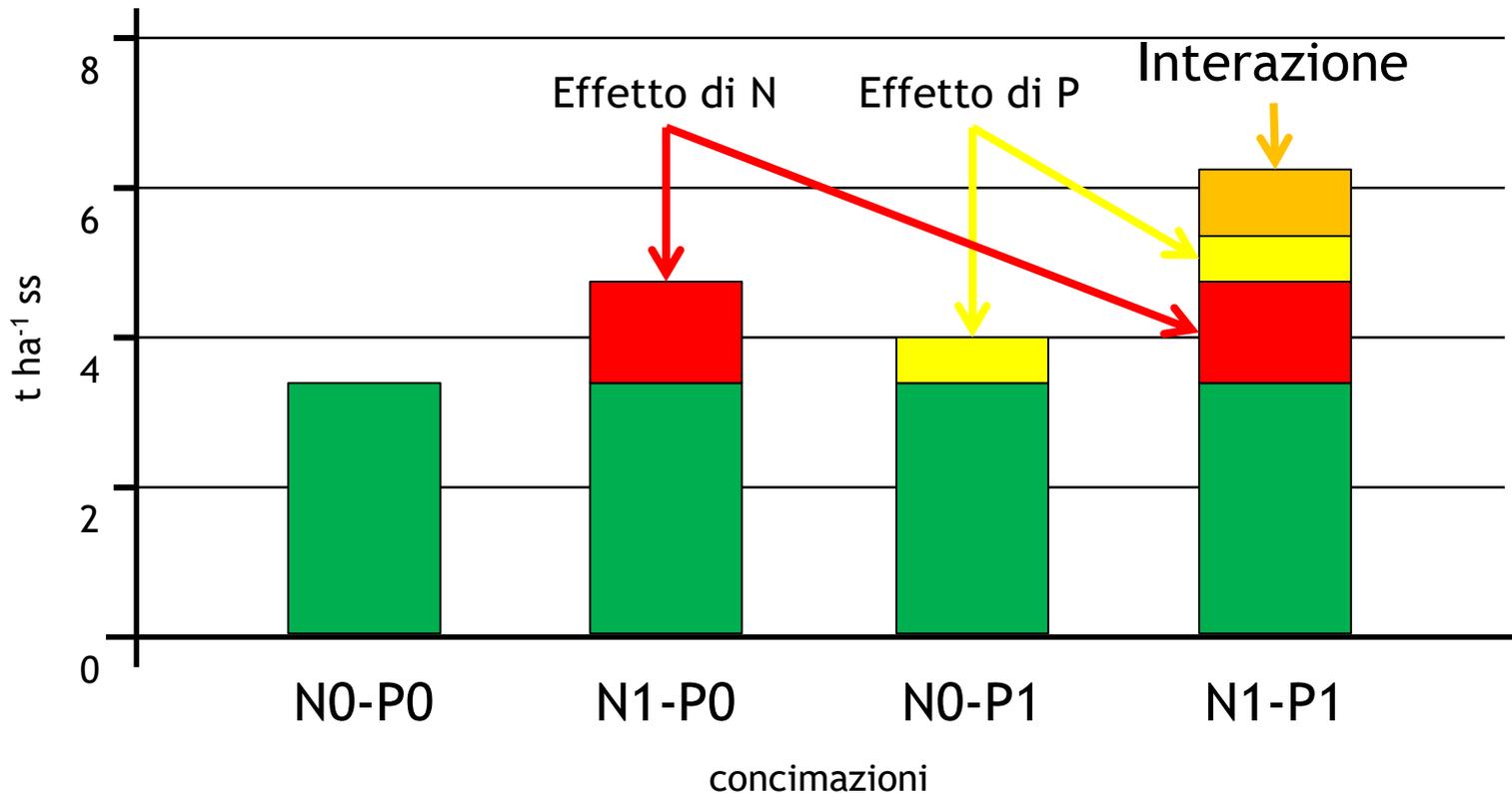
Obiettivo della sperimentazione è quello di valutare l'effetto contemporaneo dei due fattori sulla produzione di frumento.



Analisi della Varianza - Il caso a 2 o più vie

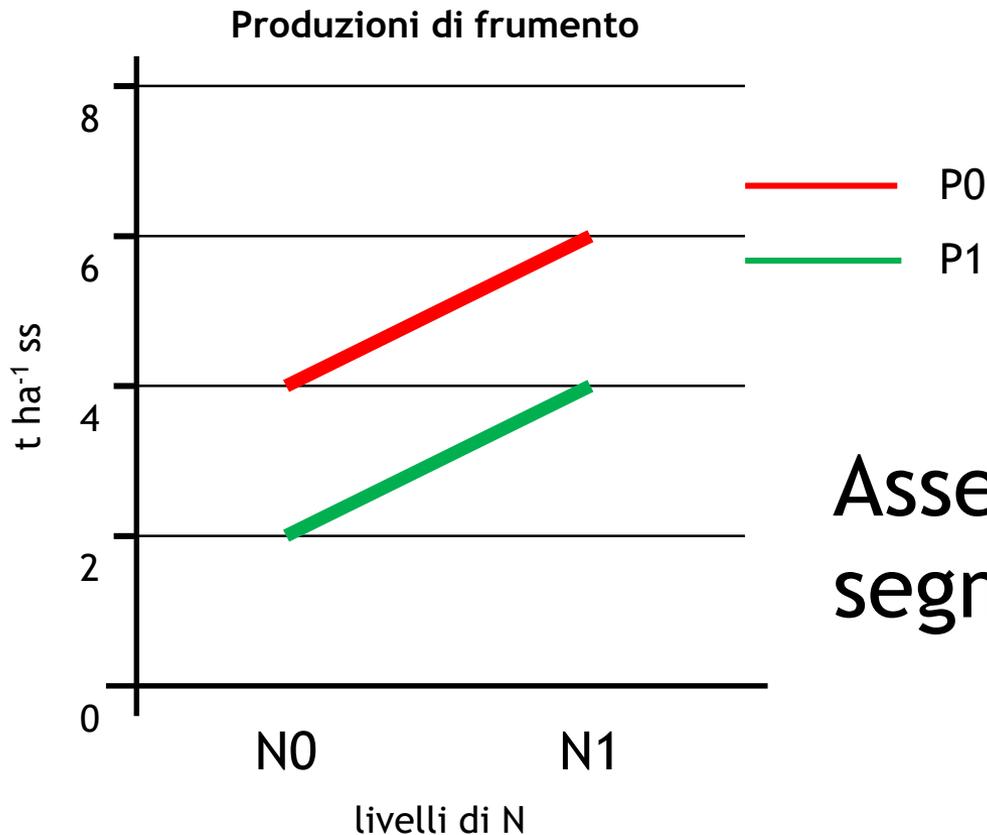
Esempio di Interazione (2/2)

Produzioni di frumento



Analisi della Varianza - Il caso a 2 o più vie

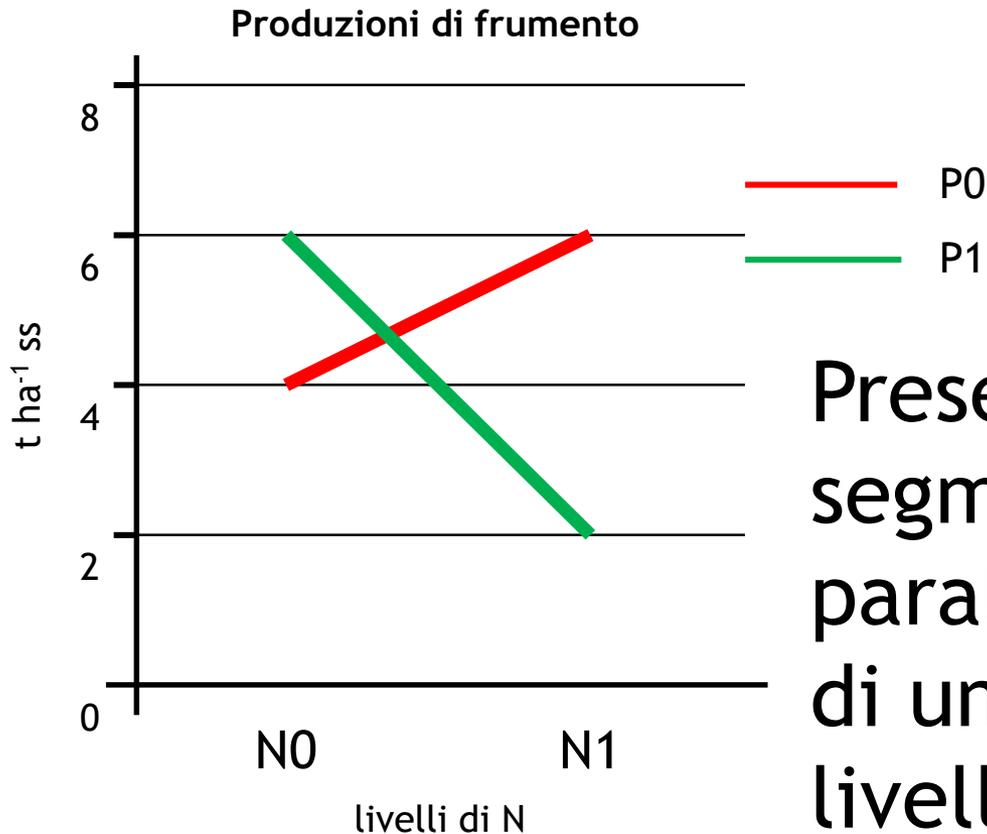
Visualizzazione dell'Interazione (1/2)



Assenza di interazione: i segmenti sono paralleli!

Analisi della Varianza - Il caso a 2 o più vie

Visualizzazione dell'Interazione (2/2)



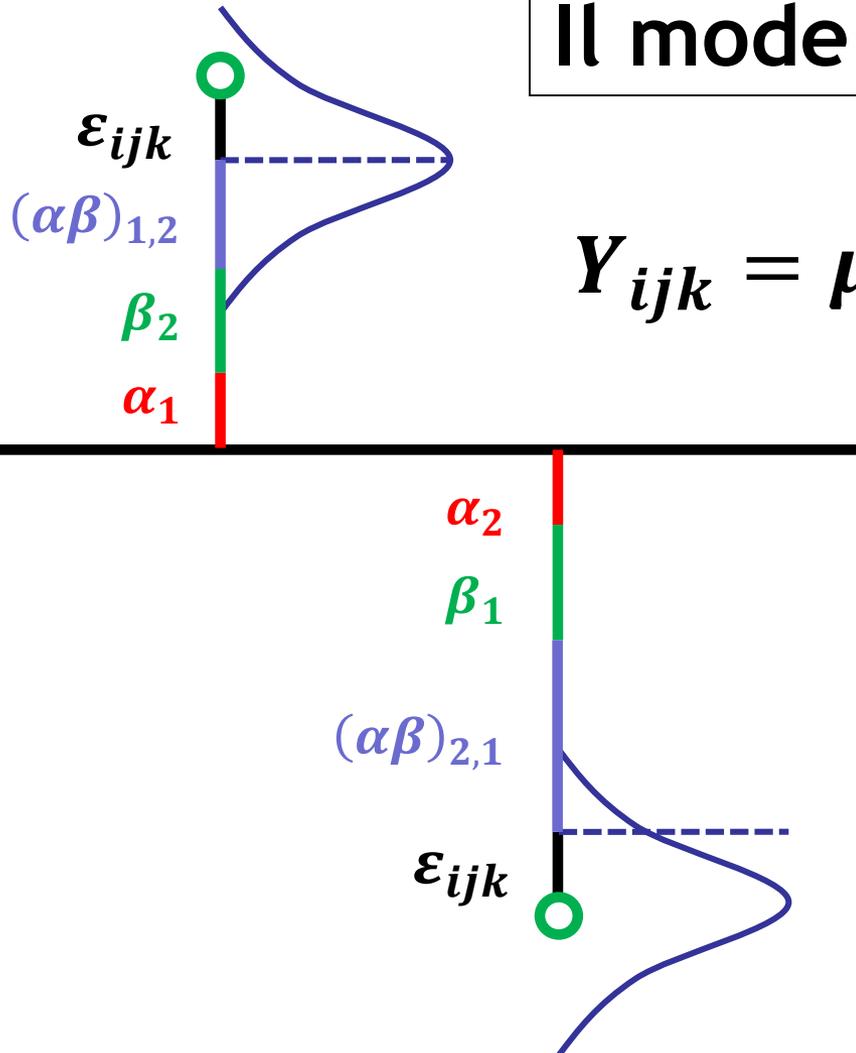
Presenza di interazione: i segmenti NON sono paralleli, poiché l'effetto di un fattore dipende dal livello dell'altro.

Analisi della Varianza - Il caso a 2 o più vie

Il modello a 2 vie

$$Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}$$

Il valore di un dato Y_{ijk} è la somma dell'effetto di uno specifico livello del 1° fattore (α_i), dell'effetto di uno specifico livello del 2° fattore (β_j), della loro interazione ($(\alpha\beta)_{ij}$) e di una componente accidentale (ε_{ijk}), tenuto conto della media generale μ .



Analisi della Varianza - Il caso a 2 o più vie

Generalizzazione

Gli effetti dovuti a un singolo fattore sono anche detti effetti semplici o di primo ordine, quelli dovuti all'interazione tra 2 fattori sono detti effetti di secondo ordine, quelli dovuti all'interazione tra 3 fattori sono detti effetti di terzo ordine e via discorrendo.

Al crescere del numero dei fattori, il modello matematico dell'Anova Fattoriale diventa subito molto complesso:

con k fattori occorre infatti sommare, oltre alla media generale e alla componente accidentale, k effetti semplici, $\binom{k}{2}$ effetti del 2° ordine, $\binom{k}{3}$ effetti del 3° ordine, ..., $\binom{k}{k-1} = k$ effetti del $(k-1)$ -esimo ordine, $\binom{k}{k} = 1$ effetto del k -esimo ordine.

Con 4 fattori, parliamo già di 17 addendi...

Quando il modello tiene conto di tutti questi elementi, parliamo di modello fattoriale completo.



Analisi della Varianza - Il caso a 2 o più vie

Osservazioni

- ✓ Normalmente ci si limita a considerare il caso di ANOVA a 2 o 3 vie, per l'impossibilità pratica di interpretare interazioni troppo complesse.
- ✓ È sempre opportuno che il numero di repliche sia uguale per ogni fattore e livello del fattore (esperimento bilanciato). In questo caso infatti si garantisce la massima potenza del test statistico e l'unicità della soluzione numerica nel caso di ANOVA a più vie.



Analisi della Varianza - Il caso a 2 o più vie

Ipotesi nulla

Consideriamo il caso di un'ANOVA a 2 vie, in cui quindi sono presenti due fattori: A (p livelli) e B (q livelli).

Nel modello fattoriale completo corrispondente a questa situazione si possono verificare 3 ipotesi distinte:

- 1) nessuna differenza tra le medie di A

$$H_0: \mu_1 = \dots = \mu_p$$

- 2) nessuna differenza tra le medie di B

$$H_0: \mu_1 = \dots = \mu_q$$

- 3) nessuna interazione tra i fattori A e B

$$H_0: AB_{ij} = 0, \forall i, j$$



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (1/6)

Ipotizziamo di avere un esperimento bilanciato con r repliche.

Passo 1: stimare la devianza errore e la devianza dovuta ai trattamenti

eseguimo gli stessi calcoli necessari per l'ANOVA a 1 via, considerando le combinazioni come fossero singoli trattamenti

$$dev_{TOT} = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (x_{ijk} - \bar{\bar{X}})^2$$

$$dev_{ERR} = \sum_{i=1}^p \sum_{j=1}^q \sum_{k=1}^r (x_{ijk} - \bar{X}_{ij})^2$$

$$dev_{TRA} = dev_{TOT} - dev_{ERR}$$

N.B. $\bar{\bar{X}}$ è la **media generale**, mentre \bar{X}_{ij} è la media calcolata in corrispondenza della **combinazione dell'i-esimo livello del trattamento A con il j-esimo livello del trattamento B**.



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (2/6)

una volta ottenute la devianza dei trattamenti nel loro complesso dev_{TRA} (utile successivamente) e la devianza errore dev_{ERR} , calcoliamo i gdl corrispondenti e, di conseguenza, le rispettive varianze

$$gdl_{TOT} = pqr - 1 \longrightarrow s^2_{TOT} = \frac{dev_{TOT}}{pqr - 1}$$

$$gdl_{ERR} = pqr - pq \longrightarrow s^2_{ERR} = \frac{dev_{ERR}}{pqr - pq}$$

$$gdl_{TRA} = gdl_{TOT} - gdl_{ERR} = pq - 1 \longrightarrow s^2_{TRA} = \frac{dev_{TRA}}{pq - 1}$$



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (3/6)

Passo 2: stimare le varianze dovute agli effetti semplici

consideriamo l'esperimento come se fosse presente un solo fattore (A)

si hanno quindi rq unità sperimentali che hanno ricevuto un determinato livello del trattamento A

per calcolare la devianza del fattore A è allora sufficiente sostituire ogni dato con la media del livello di A cui appartiene e calcolare la devianza di tutti i dati così ottenuti:

$$dev_A = rq \sum_{i=1}^p (\bar{X}_i - \bar{\bar{X}})^2 \longrightarrow s^2_A = \frac{dev_A}{p - 1}$$

In maniera del tutto analoga si procede per il fattore B



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (4/6)

Passo 3: stimare la varianze dovuta all'interazione

la via più semplice è per differenza:

$$\left\{ \begin{array}{l} dev_{INT} = dev_{TRA} - dev_A - dev_B \\ gdl_{INT} = gdl_{TRA} - gdl_A - gdl_B \\ s^2_{INT} = \frac{dev_{INT}}{gdl_{INT}} \end{array} \right.$$

dev_{INT} si può anche calcolare indipendentemente (il che può essere utile nel caso di ANOVA a 3 o più vie):

1. ponendo ogni dato uguale al valore medio della combinazione cui appartiene,
2. sottraendo a ogni dato i valori medi del livello del fattore cui si riferisce,
3. calcolando la devianza dei dati così ottenuti.

N.B. $gdl_{INT} = pq - 1 - (p - 1) - (q - 1) = (p - 1)(q - 1)$



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (5/6)

Passo 4: impostare la tabella dell'ANOVA

Fonti di variazione	Devianze	gdl	Varianze	F calcolato	P(F)
Totale	dev_{TOT}	$pqr - 1$			
Trattamenti	dev_{TRA}	$pq - 1$	S^2_{TRA}	S^2_{TRA} / S^2_{ERR}	
Fattore A	dev_A	$p - 1$	S^2_A	S^2_A / S^2_{ERR}	
Fattore B	dev_B	$q - 1$	S^2_B	S^2_B / S^2_{ERR}	
Interazione AxB	dev_{INT}	$(p - 1)(q - 1)$	S^2_{INT}	S^2_{INT} / S^2_{ERR}	
Errore	dev_{ERR}	$pqr - pq$	S^2_{ERR}		

N.B. Se A e B sono fattori fissi, il test di tutti gli effetti si fa contro l'errore!



Analisi della Varianza - Il caso a 2 o più vie

Esecuzione dei calcoli (6/6)

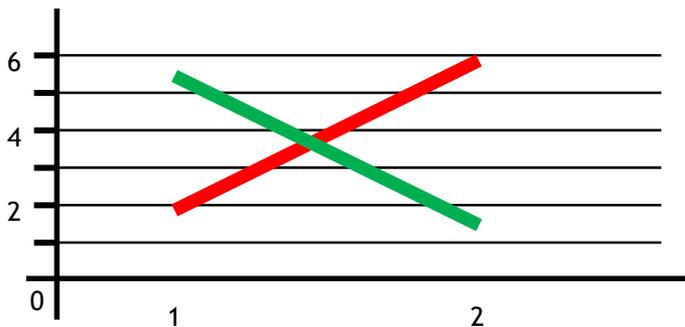
Passo 5: interpretazione

Occorre osservare innanzitutto la significatività dell'interazione: se l'effetto interattivo è significativo, allora va considerata solo l'interazione e **NON È LECITA ALCUNA CONCLUSIONE SUGLI EFFETTI SEMPLICI**. Infatti in presenza di interazione, l'effetto di un fattore è condizionato dal livello dell'altro.

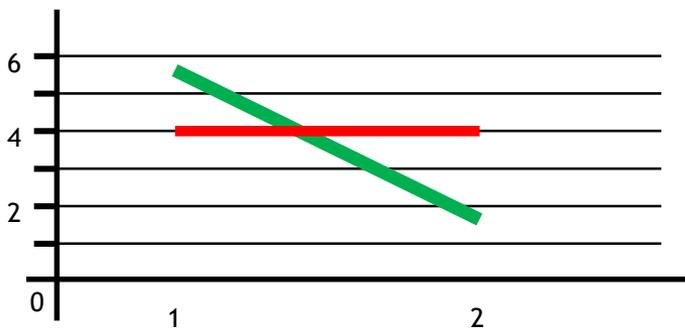


Analisi della Varianza - Il caso a 2 o più vie

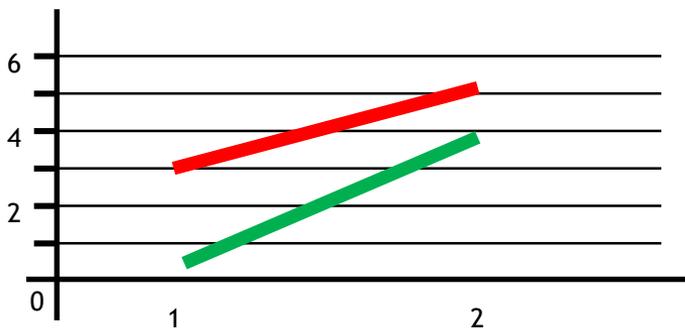
Esempi



Interazione e 2 effetti semplici significativi.



Interazione e 1 solo effetto semplice significativo.



Interazione e effetti semplici significativi. L'interazione, in questo caso, può avere valore pratico trascurabile.



Analisi della Varianza - Il caso a 2 o più vie

Esempio di calcolo (1/4)

Consideriamo l'esempio (didattico) presentato in precedenza.

Dataset

	r1	r2	r3	media
N0-P0	4,0	4,9	4,3	4,4
N0-P1	4,2	5,5	5,9	5,2
N1-P0	5,2	5,2	5,8	5,4
N1-P1	6,6	7,2	7,8	7,2

sostituiamo ogni dato con la media

Calcolo dev_{TRA}

	r1	r2	r3
N0-P0	4,4	4,4	4,4
N0-P1	5,2	5,2	5,2
N1-P0	5,4	5,4	5,4
N1-P1	7,2	7,2	7,2

$$dev_{TRA} = 12,57$$

$$gdl_{TRA} = 3$$

$$dev_{TOT} = 15,53$$

$$gdl_{TOT} = 11$$

$$dev_{ERR} = 15,53 - 12,57 = 2,96$$

$$gdl_{ERR} = 11 - 3 = 8$$

$$s^2_{ERR} = 2,96/8 = 0,37$$



Analisi della Varianza - Il caso a 2 o più vie

Esempio di calcolo (2/4)

Calcolo della devianza dovuta all'effetto N (ogni dato è sostituito con la media del corrispondente livello di N)

N0	4,8	4,8	4,8	4,8	4,8	4,8
N1	6,3	6,3	6,3	6,3	6,3	6,3

Devianza N = 6,75
gdl=1

Tabella delle medie

	P0	P1	media
N0	4,4	5,2	4,8
N1	5,4	7,2	6,3
Media	4,9	6,2	

Calcolo della devianza dovuta all'effetto P (ogni dato è sostituito con la media del corrispondente livello di P)

P0	4,9	4,9	4,9	4,9	4,9	4,9
P1	6,2	6,2	6,2	6,2	6,2	6,2

Devianza P = 5,07
gdl=1



Analisi della Varianza - Il caso a 2 o più vie

Esempio di calcolo (3/4)

Calcolo della devianza dovuta all'interazione

$$dev_{INT} = dev_{TRA} - dev_N - dev_P$$

$$gdl_{INT} = gdl_{TRA} - gdl_N - gdl_P$$

oppure

	r1	r2	r3
N0-P0	4,4	4,4	4,4
N0-P1	5,2	5,2	5,2
N1-P0	5,4	5,4	5,4
N1-P1	7,2	7,2	7,2

-

	r1	r2	r3
N0	4,8	4,8	4,8
N0	4,8	4,8	4,8
N1	6,3	6,3	6,3
N1	6,3	6,3	6,3

-

	r1	r2	r3
P0	4,9	4,9	4,9
P1	6,2	6,2	6,2
P0	4,9	4,9	4,9
P1	6,2	6,2	6,2

=

=

-5,3	-5,3	-5,3
-5,8	-5,8	-5,8
-5,8	-5,8	-5,8
-5,3	-5,3	-5,3



$$dev_{INT} = 0,75$$

$$gdl_{INT} = 1$$

Analisi della Varianza - Il caso a 2 o più vie

Esempio di calcolo (4/4)

Tabella ANOVA

Fonti di variazione	Devianze	gdl	Varianze	F calcolato	P(F)
Totale	15,5	11			
Trattamenti	12,6	3	4,9	11,32	0,003
Fattore N	6,8	1	6,75	18,24	0,003
Fattore P	5,1	1	5,07	13,70	0,008
Interazione NxP	0,8	1	0,75	2,03	0,192
Errore	2,96	8	0,37		

Nonostante un'apparente divergenza, i due segmenti sono da considerarsi paralleli in senso statistico: $P(F_{INT}) = 0,2$ non fornisce evidenze sulla presenza di interazione. Al contrario, l'effetto migliorativo del P e dell'N è significativo.

