



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 06 - Sommario

- ❑ Fattori Fissi e Fattori Casuali
- ❑ Il caso a 1 via «random»
- ❑ Modelli Misti



Definizione (1/2)

Un corretto disegno sperimentale deve garantire che la variabilità non controllata (quella non associata alle ipotesi in esame) sia equamente distribuita tra i fattori confrontati, in modo che i loro effetti possano essere separati da altre sorgenti di variazione.



Il fattore è una variabile predittiva (indipendente) della variabile di risposta (dipendente) che misuriamo. Esso è rappresentato da almeno 2 livelli, ciascuno dei quali viene esposto alle stesse condizioni sperimentali.

Definizione (2/2)

Un fattore si dice fisso quando i livelli da includere nell'esperimento sono tutti definiti dall'ipotesi e fissati dal ricercatore. Quindi tutti i livelli rilevanti per l'ipotesi sono inclusi nell'esperimento.

Un fattore si dice casuale o random quando solo una parte dei possibili (infiniti) livelli, che verranno scelti in modo casuale, sono inclusi nell'esperimento. In altre parole i trattamenti sono un campione di infiniti possibili trattamenti e non interessa studiarli in quanto tali, ma per identificare se costituiscono una fonte di variabilità significativa.

Analisi della Varianza - Fattori Fissi e Fattori Casuali

Differenza concettuale

La natura non fornisce fattori etichettati come fissi e fattori etichettati come random: ogni fattore può assumere l'una o l'altra caratteristica in funzione dell'ipotesi da analizzare e la distinzione è basata sul modo in cui i livelli del fattore sono scelti.

→ Immaginiamo di voler confrontare l'effetto di 4 fitofarmaci allo scopo di determinare quale tra questi sia il più efficace nel controllo di una certa fitopatia. In questo caso parliamo di fattore fisso con 4 livelli.

→ Immaginiamo ora di voler studiare in generale l'effetto dell'utilizzo dei fitofarmaci selettivi nel controllo della stessa fitopatia dell'esempio precedente. Dal momento che è impossibile valutarli tutti, occorre selezionarne casualmente un certo numero e di conseguenza dobbiamo parlare di fattore casuale.



Vantaggi & Svantaggi

Il fattore fisso permette di valutare gli effetti del trattamento senza errore, ma le conclusioni devono essere limitate ai soli livelli esaminati.

Il fattore casuale (o random) permette di valutare gli effetti del trattamento con errore. Esiste cioè una varianza associata a ciascun effetto, perché l'esperimento include solo un campione dei possibili (infiniti) livelli. Tuttavia, le conclusioni sono generalizzabili a tutto l'insieme da cui i livelli sono stati estratti.

Analisi della Varianza - Fattori Fissi e Fattori Casuali

Esempi di fattori casuali

- ✓ gli anni nei quali è ripetuta una sperimentazione, poiché non si è interessati agli specifici anni, ma li si considera come un campione di possibili andamenti meteorologici;
- ✓ le località, se intese solo come possibili campi in cui si ha una certa coltura e NON come rappresentative di areali;
- ✓ la variabilità delle popolazioni di partenza per ottenere incroci in genetica, per rispondere alla domanda: «ha effetto la variabilità genetica del genitore maschile, di quello femminile o la loro interazione nel determinare il risultato dell'incrocio?»

N.B. In agronomia sono frequenti esperimenti con 1 fattore random e altri fattori fissi, poco frequenti quelli con più di 1 fattore random e altri fissi, rarissimi quelli con tutti i fattori random.



Il modello a 1 effetto casuale

Immaginiamo di voler indagare l'effetto di 1 singolo fattore random. Ne consegue che il modello lineare dev'essere simile a quello dell'Anova a 1 via con fattore fisso:

$$Y_{ij} = \mu + T_i + \varepsilon_{ij}$$

ma

in questo caso l'effetto del trattamento (T_i) è a sua volta una variabile casuale. Infatti ad ogni ripetizione del medesimo esperimento, i livelli del trattamento cambiano.

Abbiamo quindi l'ipotesi aggiuntiva $(T_i) \sim N(0; \sigma_T^2)$.

Analisi della Varianza - Il caso a 1 via «random»

Conseguenze

- ✓ Le fonti di variabilità casuale (o di errore) sono diventate 2.
- ✓ Le osservazioni, all'interno di ciascun gruppo, sono correlate (cioè non indipendenti).
- ✓ La variabilità fra i gruppi non è un effetto dovuto a un trattamento, ma una semplice componente aggiunta di varianza fra i gruppi, di cui non siamo in grado di precisare esattamente le cause, se non riferendoci in modo generico a differenze ambientali imprecisate che inevitabilmente vengono a formarsi.
- ✓ **$H_0: \sigma_T^2 = 0$**
- ✓ In linea di principio la normalità del fattore casuale dovrebbe essere verificata.
- ✓ Non ha senso fare confronti post hoc sui fattori casuali.



Introduzione (1/2)

- ✓ La presenza di fattori casuali non modifica i calcoli di devianze, varianze e gradi di libertà. Sono diversi invece i test F nel caso di Modelli Misti o Mixed Models, che vedono la presenza contemporanea di fattori fissi e casuali, o di modelli completamente random (come nel caso precedente).
- ✓ Qui viene data una spiegazione intuitiva della diversa struttura dei test F; per un approccio matematico formale si rimanda a testi specifici (es. Snedecor e Cochran, 1989).
- ✓ I package di calcolo statistico hanno da non molti anni introdotto il trattamento completo dei fattori random, tuttavia occorre riconoscere un fattore random in quanto tale.



Introduzione (2/2)

- ✓ In generale, in presenza di un fattore random (β_j) nella stima della varianza di un effetto relativo a un fattore fisso (α_i) è sempre compresa una quota di varianza attribuibile al fattore casuale: infatti la stima degli α_i è fatta in funzione dei β_j , i quali sono soggetti a variazione campionaria.
- ✓ Nella stima della varianza di un fattore random non è invece compresa una quota di varianza legata al fattore fisso, dal momento che gli α_i sono sempre gli stessi.
- ✓ L'interazione tra un fattore fisso e un fattore casuale è a sua volta un fattore casuale, essendo attesi effetti diversi se l'esperimento viene ripetuto.



Analisi della Varianza - Modelli Misti

Esempio: 1 Fisso + 1 Random

Il più semplice modello misto che si può immaginare è naturalmente:

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

dove α è fisso con p livelli, β è random con q livelli e l'esperimento è bilanciato con n repliche.

Si può dimostrare che in questo caso si ha:

$$E(\text{Var}(\alpha)) = \sigma^2 + n\sigma_{\alpha\beta}^2 + nq \frac{\sum_{i=1}^p \alpha_i^2}{p-1}$$

$$E(\text{Var}(\beta)) = \sigma^2 + np\sigma_{\beta}^2$$

$$E(\text{Var}(\alpha \times \beta)) = \sigma^2 + n\sigma_{\alpha\beta}^2$$

$$E(\text{Var}(\text{errore})) = \sigma^2$$



Osservazioni (1/2)

- ✓ L'effetto del fattore fisso α si valuta testando la sua varianza contro la varianza dell'interazione $\alpha \times \beta$.
- ✓ L'effetto del fattore casuale β si valuta testando la sua varianza contro la varianza dell'errore σ^2 .
- ✓ L'effetto dell'interazione $\alpha \times \beta$ si valuta testando la sua varianza contro la varianza dell'errore σ^2 .

In generale, in presenza di 1 solo fattore random, ogni effetto fisso deve essere testato contro la sua interazione con il fattore random, mentre il fattore random e le interazioni tra ogni fattore fisso e il fattore random devono essere testate contro l'errore.

Osservazioni (2/2)

*Intuitivamente è abbastanza chiaro che, per dichiarare un fattore fisso «significativo» in presenza di fattori casuali, il suo effetto deve essere **superiore** non solo a quello della **variabilità sperimentale**, ma anche a quello della **variabilità indotta dal fattore casuale**. In questo modo infatti è possibile garantire che esso è «apprezzabile» a prescindere dagli eventuali altri valori che il fattore casuale potrebbe assumere, al di là di quelli analizzati.*

*Questo è il motivo per il quale il test corretto da utilizzare per valutare un fattore fisso è quello contro la sua interazione con il fattore casuale, il quale comprende sia una quota dovuta alla **variabilità accidentale** (come ogni varianza stimata), sia una quota di **interazione tra il fattore fisso e il fattore random**.*



Analisi della Varianza - Modelli Misti

Esempio

Immaginiamo di aver condotto un esperimento per valutare la produttività di 3 vitigni. Per ciascun vitigno sono stati rilevati i dati di resa di un certo numero di vigneti (considerati rappresentativi) e la procedura è stata ripetuta per 5 anni di seguito, sorteggiando ogni volta vigneti diversi.

Siamo chiaramente in presenza di uno schema sperimentale con un fattore fisso a 3 livelli (il vitigno) e un fattore random a 5 livelli (l'anno).

Vediamo ora che conclusioni possiamo trarre dall'analisi della varianza:

se l'interazione è significativa allora **STATISTICAMENTE** non possiamo dire nulla a proposito della produttività dei 3 vitigni;

se l'interazione non è significativa, allora possiamo trarre delle conclusioni generali a proposito dei 3 vitigni, guardando la significatività dell'effetto semplice.

