

# Esperimenti multi-ambiente (MET)

Andrea Onofri

Perugia, February 2011

## Contents

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Esempio</b>	<b>3</b>
<b>3</b>	<b>ANOVA combinata</b>	<b>4</b>
<b>4</b>	<b>Stabilità e affidabilità produttiva</b>	<b>7</b>

## 1 Introduzione

### Introduzione

- *All research should cover sufficient temporal and spatial variation to be able to make sound generalisations. Such a statement does not only refer to field and glasshouse trials, but also to experiments carried out in more controlled conditions, with material that is subject to spatial and temporal variability (seed populations, batches of treated soils, etc.).* (from: Weed Resaerch, guidelines for authors)
- Tutti gli esperimenti debbono essere ripetuti in più ambienti
- In questo caso, è necessaria un'analisi finale cumulativa
- Se l'interazione 'trattamento x ambiente' è significativa, il risultato di esperimenti multipli è sotto il controllo ambientale
- Non ha senso parlare di risultati medi!

### Prove varietali

- Le prove varietali sono decisamente sotto il controllo ambientale e, pertanto, debbono sempre essere ripetute, se si vogliono dare raccomandazioni attendibili agli agricoltori

- Ambienti = anni e/o località
- Le raccomandazioni varietali debbono considerare i cosiddetti macro-areali, nei quali le varietà danno prestazioni simili
- Oltre ad informazioni produttive, le raccomandazioni varietali debbono considerare anche la stabilità produttiva

### **Reti varietali**

- Reti sperimentali con più unità operative, che provano le stesse varietà in più anni e più località, utilizzando lo stesso protocollo (blocco randomizzato con 3-4 repliche)
- Oltre al gruppo di varietà comuni, in ogni località vengono anche incluse cultivars di interesse locale
- Nel corso degli anni, le varietà poco interessanti vengono abbandonate, mentre vengono inserite in prova quelle di nuova costituzione
- E' fondamentale mantenere comunque un gruppo di varietà costantemente presenti in tutti gli anni e tutte le località

### **Obiettivi delle reti varietali**

- Valutare le migliori varietà in ogni ambiente (o gruppi di ambienti affini)
- Valutare le migliori varietà in assoluto
- Valutare la presenza di interazioni 'crossover' tra ambienti e varietà (In altre parole, valutare se la classifica varietale cambia negli ambienti)
- Per ogni ambiente, valutare la produttività media attesa per ogni varietà
- Per ogni ambiente, valutare la stabilità produttiva interannuale

### **Gli ambienti come super-repliche (fattore random?)**

- In molti casi, le località e, soprattutto, gli anni non sono interessanti di per se', ma possono essere considerati un campione rappresentativo di un universo più ampio di ambienti possibili
- In questo caso gli ambienti agiscono da super-repliche e vengono definiti due livelli di variabilità: (1) entro prova (2) tra prove. Queste due fonti di variabilità debbono essere valutate separatamente

### **Quando è possibile considerare gli ambienti 'random'**

- Anni e/o ambienti possono essere considerati 'random' se (1) sono sufficientemente numerosi e (2) sono rappresentativi
- Quest'ultimo requisito è di solito soddisfatto per gli anni, ma non è sempre soddisfatto per le località, in quanto
  - Le località non sono scelte casualmente, ma sono quelle dove si trovano stazioni sperimentali appositamente attrezzate
  - Spesso vi è un effettivo interesse locale specifico
- LA SCELTA DI CONSIDERARE GLI AMBIENTI COME FATTORE FISSO O RANDOM E' CONTESTUALE E STA NELLE MANI DELLO SPERIMENTATORE

## **2 Esempio**

### **Progetto RinnovaProve - 2009**

- 12 località Italiane (da Nord a Sud)
- 17 varietà di pisello proteico
- Blocchi randomizzati con quattro repliche in ogni località
- dataset a: [www.casaonofri.it/Biometry/Presentations/pisello.xls](http://www.casaonofri.it/Biometry/Presentations/pisello.xls)

### **Analisi Preliminari**

- Elaborazioni separate (ANOVA e MCP)
- Domandarsi se tutte le località/anni forniscono dataset di buona qualità
- Valutare se gli errori sperimentali delle diverse prove sono omogenei
- Eseguire il 'F Max' test per l'omogeneità delle varianze
- Se le varianze dei diversi ambienti non sono omogenei, divider gli ambienti in sottogruppi con varianze più o meno omogenee
- Procedere all'elaborazione cumulata delle prove

### Test 'F max'

- Rapporto tra la varianza residua dell'ambiente a maggior variabilità con quella dell'ambiente a minor variabilità. In questo esempio:

$$F_{MAX} = \frac{0.5352}{0.0231} = 23.17$$

- Considerando i gradi di libertà (in questo caso 48 sia al numeratore che al denominatore) il test è altamente significativo
- CONCLUSIONE: Le varianze di prove diverse non sono omogenee

### Cosa fare per varianze non-omogenee?

- TRASFORMARE I DATI. Ciò altera l'unità di misura e complica l'interpretazione
- PESARE I DATI. Si potrebbe assegnare alle prove con minore variabilità. Dato che le prove più variabili sono spesso quelle con produzioni medie più alte, il rischio è quello di sottovalutare proprio gli ambienti migliori!
- SUDDIVIDERE GLI AMBIENTI IN SOTTOGRUPPI. Soluzione migliore
- NON FARE NULLA. Considerare che la mancanza di omogeneità delle varianze comporta una 'diminuzione di significatività' che rende il nostro test meno potente.

## 3 ANOVA combinata

### ANOVA combinata - Ambienti Fissi

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Loc	11	1222.19	111.108	591.4861	< 2.2e-16 ***
Loc:factor(Block)	36	15.55	0.432	2.2997	3.905e-05 ***
Var	16	128.50	8.031	42.7553	< 2.2e-16 ***
Loc:Var	176	240.36	1.366	7.2703	< 2.2e-16 ***
Residuals	576	108.20	0.188		

Notare che l'effetto dei blocchi è considerato entro ciascun ambiente (4 blocchi, cioè 3 gradi di libertà per 12 ambienti = 36 gradi di libertà totali).

### Interpretazione - Ambienti FISSI

- L'interazione 'genotipo x ambiente' (GE) è un effetto fisso ed è altamente significativo (come quasi sempre accade).
- Di conseguenza, le varietà si sono comportate in modo diverso nei diversi ambienti e possono essere degli effetti 'cross-over'.

- Non ha nessun senso considerare le medie varietali.
- L'analisi cumulativa serve quindi a poco, in termini di raccomandazione varietale.

### ANOVA combinata - Ambienti Random

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Loc	11	1222.19	111.108		
Loc:factor(Block)	36	15.55	0.432		
Var	16	128.50	8.031	5.8808	3.443e-10 ***
Loc:Var	176	240.36	1.366		
Residuals	576	108.20	0.188		

### Interpretazione: ambienti sono random

- Oltre agli ambienti sono random anche tutte le interazioni con l'ambiente, inclusa l'interazione GE
- Le varietà si comportano in modo diverso in ogni ambiente, ma questo non ci interessa, perchè gli ambienti fungono solo da 'super-repliche'.
- Posso considerare le medie varietali, debbo però considerare che su queste agiscono più fonti di variabilità
- SED, SEM e confronti multipli debbono essere basati su tutte le fonti di variabilità che li influenzano (componenti di varianza).

### Variance components degli effetti 'random'

- Le componenti di varianza rappresentano le varianze produttive dovute all'ambiente ( $\sigma_e$ ), ai blocchi entro ambiente ( $\sigma_{e|b}$ ), all'interazione 'varietà x ambiente' ( $\sigma_{GE}$ ) e all'errore sperimentale entro prova ( $\sigma_{res}$ ).
- Sono varianze e, in quanto tali, la loro unità di misura è il quadrato dell'unità originale dei dati.
- Secondo il metodo dei momenti, le componenti di varianza si calcolano a partire dalle varianze nell'ANOVA (MSE).
- Le varianze dell'ANOVA debbono essere depurate dall'errore sperimentale (che è sempre presente) e dalle altre fonti di variabilità legate ai fattori 'random'.
- Le spiegazioni sulle metodiche di calcolo si trascurano, assumendo che lo studente si affidi ad un software statistico

### Esempio del calcolo delle 'variance components'

$$\sigma_{res}^2 = MSE_{res} = 0.188$$

$$\sigma_{ge}^2 = \frac{MSE_{ge} - \sigma_{res}^2}{b} = \frac{1.366 - 0.188}{4} = 0.294$$

$$\sigma_{b|e}^2 = \frac{MSE_{b|e} - \sigma_{res}^2}{g} = \frac{0.432 - 0.188}{17} = 0.014$$

$$\sigma_e^2 = \frac{MSE_e - g\sigma_{b|e}^2 - r\sigma_{ge}^2 - \sigma_{res}^2}{rg} = \frac{111.11 - 17 \times 0.014 - 4 \times 0.294 - 0.188}{4 \times 17} = 1.610$$

I valori di MSE rappresentano le voci di varianza nell'ANOVA,  $b$  è il numero dei blocchi,  $g$  è il numero delle varietà ed  $e$  è il numero degli ambienti

### SEM e SED

- I SEM e i SED risentono non solo della variabilità entro prova (residuo), ma anche di tutte le altre fonti di variabilità (effetti random).
- Le medie produttive risentono di  $\sigma_e$ ,  $\sigma_{b|e}$ ,  $\sigma_{ge}$  e di  $\sigma_{res}$ .
- Le differenze produttive risentono solo  $\sigma_{ge}$  e  $\sigma_{res}$ . Ambiente e blocco agiscono uniformemente per tutte le varietà e quindi non ne modificano le differenze.

$$SED = \sqrt{2 \times \frac{r\sigma_{ge} + \sigma_{res}}{re}} = \sqrt{2 \times \frac{4 \times 0.294 + 0.188}{48}} = 0.2385$$

### Valutazione medie produttive

- In questo caso, si può utilizzare un test di confronto multiplo con il migliore (MCB; test di Dunnet ad una coda)
- La differenza critica è: 0.623
- La miglior varietà è Attika, in un gruppo di 8 varietà di vertice

Attika	4.02	a
Gregor	4.00	a
Arthur	3.76	a
Standal	3.68	a
Starter	3.64	a
Enduro	3.44	a
Prelude	3.43	a
Pactol	3.41	a
Pepone	3.35	
Corrent	3.29	
Isard	3.28	
Belot	3.27	
Amical	3.27	
Royal	3.23	
Javlo	3.15	
Cartouche	2.98	
Simbol	2.25	

## 4 Stabilità e affidabilità produttiva

### Stabilità e affidabilità produttiva

- La classifica media in termini di produzione è un buon elemento per la raccomandazione varietale, ma non è sufficiente
- Infatti, soprattutto quando abbiamo molte località, potrebbero essere presenti interazioni 'crossover'
- La classifica (e quindi la raccomandazione varietale) cambia in funzione dell'ambiente
- Nasce l'esigenza di capire se le varietà, oltre che produttive sono anche 'stabili'
- E' necessario guardare con più attenzione l'interazione 'genotipo ambiente'

### Medie produttive e deviazioni standard

- Forniscono previsioni produttive direttamente utilizzabili
- La deviazione standard delle medie varietali è una misura di **stabilità statica** (environmental 'variance': più è bassa più la produzione di un genotipo rimane stabile nei diversi ambienti)
- Questo tipo di stabilità non necessariamente è positivo: implica infatti che un genotipo non è svantaggiato in ambienti sfavorevoli, ma non riesce neanche ad approfittare di condizioni favorevoli.

- La stabilità statica è piuttosto importante in agricoltura biologica, dove spesso le condizioni agronomiche sono sfavorevoli
- I dati produttivi ottenuti in diversi ambienti non sono direttamente confrontabili, se non in relazione alla media dell'ambiente

### **Indici produttivi**

- Le produzioni di ciascuna varietà vengono espresse in percentuale rispetto alla media dell'ambiente
- Varietà con valori sopra a 100 sono migliori della media, in ogni ambiente
- Con questa tabella si rimuove l'effetto dell'ambiente ed i dati produttivi ottenuti in ambienti diversi sono direttamente confrontabili
- Si perde l'unità di misura
- Le medie degli indici varietali ci danno informazioni anche sul ranking delle medie: se una varietà ha una media superiore a 100 significa che nella maggior parte degli ambienti ha produzioni sopra la media (e quindi occupa un posto alto in graduatoria)
- Le deviazioni standard sono un indice di **stabilità dinamica** (una varietà è stabile se segue le oscillazioni ambientali, in senso positivo e/o negativo)

### **Effetti GGE**

- Dalle produzioni di ciascuna varietà vengono sottratte le medie ambientali
- Varietà con valori positivi sono sopra la media
- Con questa tabella, si rimuove l'effetto dell'ambiente, conservando l'unità di misura
- Le deviazioni standard sono un indice di stabilità dinamica ('ecovalenza': una varietà è stabile se segue le oscillazioni ambientali, in senso positivo e/o negativo)

### **Stabilità e affidabilità**

- La scelta varietale dovrebbe essere fatta in base sia alla produttività media, sia alla stabilità produttiva
- Si parla quindi di 'affidabilità produttiva': è affidabile una varietà sia produttiva che stabile



- Affidabilità produttiva secondo Annichiarico:

$$I_i = Y_i - Z_P \cdot S_i$$

per la  $i$ -esima varietà,  $Y$  è l'indice produttivo medio,  $S$  è la deviazione standard tra ambienti,  $Z_P$  è la variabile normale standardizzata corrispondente a  $p = 0.75$ , cioè il 75° centile, corrispondente a  $Z_P = 0.675$ ).

- L'indice di affidabilità corrisponde alla produzione ottenibile in tre casi su quattro in un certo ambiente. Più è alta l'affidabilità più  $Y$  e  $I$  si avvicinano.

### **Altri approcci più avanzati**

- Rank sum index (approccio nonparametrico)
- Cluster analysis
- Factor analysis
- Principal component analysis
- AMMI and GGE analysis
- Pattern analysis
- Analysis of crossovers
- Partial Least Squares Regression
- Factorial Regression

### **Approfondimenti**

Annichiarico, P., 2002. Genotype x Environment Interactions - Challenges and Opportunities for Plant Breeding and Cultivar Recommendations. FAO, 1, 1-115 E' un testo on-line. Può facilmente essere reperito tramite Google.