



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 05 - Sommario

- Analisi della Varianza
 - Confronti tra medie
 - A priori
 - A posteriori



Analisi della Varianza - Confronti tra medie

Il problema

L'analisi della varianza è una tecnica che consente di operare un confronto tra 2 o più medie, partendo sempre dal presupposto (Ipotesi Nulla), che esse siano uguali.

Ne consegue che rifiutare l'Ipotesi Nulla tramite il Test F equivale ad affermare che almeno una delle medie considerate è significativamente diversa dalle altre.

Una volta giunti a questo punto però non si può dire nient'altro, dal momento che non si sa quali siano i rapporti di differenza reciproca tra le medie.



Analisi della Varianza - Confronti tra medie

La soluzione

ortogonali a priori
o pianificati

Confronti tra Medie

multipli o a posteriori
(post hoc)



Confronti ortogonali a priori (1/2)

I confronti a priori prendono il loro nome dal fatto che sono definiti in fase di pianificazione dell'attività, vale a dire prima dell'esperimento.

I confronti a priori tra le medie dovrebbero essere ortogonali, il che equivale a richiedere che siano indipendenti. In altre parole, ciascun confronto non deve dare informazioni sui risultati dell'altro.

L'ortogonalità dei confronti pianificati si verifica attraverso il metodo dei coefficienti dei contrasti (anche polinomiali).



Confronti ortogonali a priori (2/2)

Quando è possibile una pianificazione dell'esperimento, i confronti ortogonali tendenzialmente sarebbero da preferire come metodo di valutazione delle differenze tra specifiche medie.

- ✓ Sono potenti
- ✓ Usano tutti i dati
- ✓ Sono «razionalmente fondati»

Tuttavia

- ✓ Sono concettualmente più complicati da predisporre
- ✓ Sono praticamente più difficili da applicare

N.B. I confronti ortogonali a priori non vengono affrontati in questo corso.



Introduzione

- ✓ I confronti multipli si applicano dopo che l'ANOVA ha permesso di rifiutare H_0 (secondo la maggioranza degli autori...), allo scopo di individuare le differenze che hanno determinato la significatività totale.
- ✓ Tendenzialmente, le loro condizioni di validità sono le stesse dell'ANOVA, con robustezza differenziata rispetto all'allontanamento dai requisiti (l'eteroscedasticità è la condizione più grave nell'influenzare le probabilità degli errori di tipo I e II).

Tassi d'Errore (1/3)

La probabilità di commettere un errore di I tipo, cioè la probabilità α di trovare una differenza significativa quando in realtà essa non esiste, è corretta per il singolo confronto tra due medie. Questo tasso d'errore viene chiamato con termine tecnico comparison-wise o pair-wise.

Esso determina, all'aumentare del numero di confronti, un tasso d'errore per tutto l'esperimento notevolmente maggiore, che viene definito experiment-wise.



Analisi della Varianza - Confronti multipli

Tassi d'Errore (2/3)

Se effettuiamo un test tra due medie (cioè il test t di Student), impostando un livello di significatività $\alpha = 0.05$, questo valore coincide ovviamente con la probabilità p di commettere l'errore di I tipo.

In questo caso il numero di confronti è pari a 1 e perciò la probabilità di avere un numero di errori pari 0 è data da $1 - 0,05 = 0,95$.

In altre parole, con 2 sole medie (e quindi con 1 unico confronto) il tasso d'errore comparison-wise coincide con il tasso d'errore experiment-wise.



Analisi della Varianza - Confronti multipli

Tassi d'Errore (3/3)

Ipotizziamo ora di dover confrontare un numero di medie superiore a 2.

Immaginiamo, a titolo di esempio, di aver condotto un esperimento finalizzato a confrontare l'efficacia di 5 fitofarmaci e di aver conseguentemente applicato l'analisi della varianza a 1 via (con 1 fattore a 5 livelli).

Il test F ha portato a rifiutare $H_0: m_1 = m_2 = m_3 = m_4 = m_5$ e perciò ora abbiamo la necessità di «stilare una classifica» delle medie.

Per fare ciò, occorre confrontare ciascuna media con tutte le altre. Il numero dei confronti da effettuare è quindi pari a $\binom{n}{2} = \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{1*2*3*4*5}{1*2*(1*2*3)} = 10$.

Ciascun confronto ha un tasso d'errore comparison-wise pari a 0,05 e perciò la probabilità di avere un numero di errori pari a 0 considerando tutti i 10 confronti è pari a

$\overbrace{(1 - 0,05) * \dots * (1 - 0,05)}^{10 \text{ volte}} = 0,95^{10} = 0,5987$. Ne consegue che il **tasso d'errore experiment-wise** (cioè di tutto l'esperimento) è pari a **$1 - 0,5987 = 0,4013$** .



Il problema della scelta (1/2)

Nell'esempio che abbiamo appena proposto, la probabilità di commettere l'errore di I tipo relativa all'intero esperimento è del 40% e questo è naturalmente inaccettabile.

È chiaro quindi che nella scelta del test a posteriori è necessario un compromesso tra:

- tasso d'errore comparison-wise (α), collegato alla potenza del singolo test, e
- tasso d'errore experiment-wise (α_T), collegato al principio di cautela o protezione di tutta la serie dei test.

Il problema della scelta (2/2)

La ricerca dell'equilibrio tra potenza e protezione non ha prodotto una soluzione universalmente condivisa e ha determinato una molteplicità di proposte (spesso contrastanti), che variano in funzione di:

- situazioni sperimentali (gruppi bilanciati o meno, numero e organizzazione dei confronti);
- rischi relativi che il ricercatore intende correre rispetto alle probabilità α e β .

N.B. Si ricordi che α è la probabilità di rifiutare H_0 quando invece è vera, mentre β è la probabilità di accettare H_0 quando invece è falsa.



Analisi della Varianza - Confronti multipli

LSD e Duncan

Il metodo LSD (Least Significant Difference o Minima Differenza Significativa) si applica disponendo di k medie, ciascuna calcolata su n dati. Viene prima determinato l'intervallo di confidenza per una generica coppia di medie, con il test t di Student. Successivamente vengono considerate significative tutte le differenze che in valore assoluto superano la LSD.

La tecnica LSD è largamente impiegata, tuttavia è stato dimostrato che il tuo tasso di errore experiment-wise reale esplose non appena il numero di gruppi è superiore a 3.

Il test di Duncan è fondato su presupposti più o meno analoghi e anch'esso viene usato con frequenza; anche in questo caso però il valore reale di α cresce rapidamente all'aumentare del numero dei confronti.

LSD e Duncan sono perciò metodi fortemente sconsigliati!



Analisi della Varianza - Confronti multipli

Bonferroni e Sidak

Per effettuare p volte il test t di Student, mantenendo costante la probabilità totale α_T (experiment-wise), una delle prime soluzioni proposte è quella dovuta a **Bonferroni**. Questo metodo sostanzialmente consiste nell'imporre che la probabilità α di ciascun confronto (comparison-wise) sia minore o uguale di $\frac{\alpha_T}{p}$. Successivamente questa relazione è stata corretta da **Sidak**, il quale ha dimostrato che la vera relazione tra α e α_T è di tipo esponenziale:

$$\alpha = 1 - (1 - \alpha_T)^{\frac{1}{p}}$$

I test basati sulle correzioni di Bonferroni e soprattutto di Sidak sono consigliati, in particolare per l'analisi post-hoc dell'interazione tra fattori.



Analisi della Varianza - Confronti multipli

Test HSD di Tukey

Il metodo di confronti multipli più diffuso è quello proposto da Tukey nel 1949. Il test è anche chiamato *wholly significant difference test*, poiché pone attenzione all'errore experiment-wise, o *honestly significant difference test* (HSD). Questo test è stato ideato per arrivare ad un accettabile compromesso tra errore experiment-wise ed errore comparison-wise. Si tratta di un test di tipo experiment-wise, più potente dei test fondati sulle correzioni di Bonferroni o Sidak.

Il Test HSD di Tukey (con la correzione di Kramer del 1956) è consigliato per il confronto multiplo tra più medie, soprattutto nel caso in cui l'esperimento sia sbilanciato.



Test di Ryan-Einot-Gabriel-Welsch

Esistono due versioni del test di Ryan-Einot-Gabriel-Welsh: REGW-Q (che utilizza un intervallo studentizzato) e REGW-F (che utilizza un test F). Il primo è il più comune.

REGW-Q costituisce attualmente il metodo più generale per corrispondere al desiderio di massimizzare la potenza assicurando un adeguato livello di protezione.

Il Test di Ryan-Einot-Gabriel-Welsch è fortemente consigliato per il confronto multiplo tra le medie nel caso di esperimenti bilanciati. In caso contrario, è meglio usare il Test HSD di Tukey.

Test di Scheffé (con estensione di Gabriel)

Questa procedura è stata proposta da Scheffé nel 1953 e poi nel 1959; infine è stata generalizzata da Gabriel nel 1978.

Offre il vantaggio di eseguire anche confronti complessi (tra medie di due o più gruppi), ma è il meno potente ed inoltre aumenta la probabilità di commettere l'errore di tipo II.



Violazione dei requisiti (1/2)

Le condizioni di applicabilità dei test post hoc visti fin qui sono le stesse dell'Anova. Si pone perciò il problema di come agire in caso di violazioni (che hanno portato ad utilizzare metodi «alternativi» per verificare l'ipotesi nulla).

La materia è piuttosto controversa e non c'è accordo tra i vari autori su quale sia la strategia migliore da adottare. Noi ci limiteremo a fornire qualche indicazione di massima, tenendo presente che in molti casi occorre ricorrere a soluzioni *ad hoc*.



Violazione dei requisiti (2/2)

In caso di Non Normalità, in genere è possibile ricorrere ai test presentati nelle slide precedenti poiché essi sono sufficientemente robusti rispetto a questa violazione.

In caso di Non Omogeneità, la faccenda si fa più complessa... Nel caso dell'Anova a 1 via o comunque dell'analisi degli effetti semplici, solitamente si ricorre al test di Games-Howell. L'approccio Bonferroni-Sidak invece dovrebbe andar bene per l'interazione tra fattori, tuttavia è scarsamente utilizzato. In alternativa è possibile adottare un approccio detto «pseudo one way» e applicare lo stesso Games-Howell.

