Capitolo 3. L'analisi della varianza. Il problema dei confronti multipli. La soluzione drastica di Bonferroni ed il test ponderato di Student-Newman-Keuls. L'analisi della varianza applicata a disegni sperimentali complessi.

Il confronto globale tra più medie: l'analisi della varianza

Quando a scuola l'insegnante di educazione fisica organizza un piccolo torneo di calcetto tra alunni, in genere tende a formare squadre omogenee. In che modo: scegliendo a caso o, meglio ancora, mescolando il più possibile bravi, meno bravi e scarponi. Senza saperlo, il criterio adottato dall'insegnante è quello di avere **all'interno** delle squadre (gruppi) le stesse differenze presenti **tra** squadre diverse. L'analisi della varianza ci dice in questo caso se le squadre sono equilibrate.
Quando si considerano più gruppi (ad es. squadre di calcio o gruppi di animali trattati in modo diverso) è possibile scindere la variazione globale della variabile (bravura a calcio o pressione arteriosa o altro) in una componente dovuta a differenze **tra** i vari gruppi ed in una componente dovuta a differenze **entro** i gruppi. L'ipotesi nulla assume che tutti i gruppi derivino dalla stessa popolazione e che quindi la varianza calcolata tra gruppi e quella calcolata entro gruppi siano uguali in quanto stime della stessa varianza della popolazione. Quindi per l'ipotesi nulla il rapporto

$$F = \frac{\text{varianza tra gruppi}}{\text{varianza entro gruppi}} = 1$$

Invece, se i gruppi provengono da diverse popolazioni (con medie diverse) la varianza tra gruppi sarà maggiore della varianza entro gruppi, e quindi

$$F = \frac{\text{varianza tra gruppi}}{\text{varianza entro gruppi}} > 1$$

Per rifiutare l'ipotesi nulla, e quindi ritenere che esistano differenze significative tra le medie, occorre che il valore di F trovato ($F>1$) abbia meno del 5% di probabilità di provenire per caso dalla situazione ipotizzata dalla ipotesi nulla ($\alpha<5\%$). In tal senso, quanto più F è grande, tanto più si va verso il rifiuto dell'ipotesi nulla. Se si rifiuterà l'ipotesi nulla si accetterà l'ipotesi alternativa che sostiene che i gruppi non provengano tutti dalla stessa popolazione. Purtroppo, il test F non consente di precisare quale o quali gruppi differiscano dagli altri. Il test F è un test globale. Come quando, in certe trame poliziesche, si dimostra che tra un numero di persone c'è senz'altro l'assassino ma non si sa chi sia. Il test F è la premessa obbligatoria per scoprire l'assassino, come vedremo tra breve. Il test F descritto è il capostipite di una serie di importanti test riuniti sotto il nome di analisi della varianza (**ANOVA**, ANalysis Of VAriance) [nota bene: analisi della varianza, NON semplice calcolo della varianza].

Per calcolare la varianza, occorre comunque calcolare prima la somma dei quadrati o devianza. Esisterà quindi una devianza totale, una devianza tra gruppi ed una entro gruppi. Il 'magico' dell'analisi della varianza, è la relazione di decomposizione:

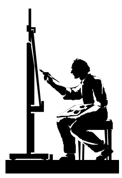
devianza totale = devianza tra gruppi + devianza entro gruppi: $S_{totale} = S_{tra\ gruppi} + S_{entro\ gruppi}$

e lo stesso per i gradi di libertà: $GDL_{totali} = GDL_{tra\ gruppi} + GDL_{entro\ gruppi}$

Per devianza totale, con i suoi gradi di libertà, si intende la devianza che stimeremmo da tutti i dati di tutti i gruppi messi assieme, come in un'unica lista. Resta solo da vedere come stimare la devianza tra gruppi e quella entro gruppi.

	TOTALE	TRA GRUPPI	ENTRO GRUPPI
DEVIANZA	mettendo insieme tutti i dati e calcolando la somma dei quadrati rispetto alla media totale	sostituendo ai dati di ciascun gruppo la media del gruppo e calcolando poi la somma dei quadrati rispetto alla media totale (in tal modo si annulla la variazione entro i gruppi)	facendo le somme dei quadrati entro ciascun gruppo e sommando insieme (in tal modo si annulla la variazione tra gruppi)
GDL	numero totale di dati - 1	numero dei gruppi - 1	somma dei GDL delle devianze di ciascun gruppo $(n_a-1) + (n_b-1) + (n_c-1) + ...$
	⇩	⇩	⇩
VARIANZA	DEVIANZA/GDL	DEVIANZA/GDL	DEVIANZA/GDL

Il test F è sempre un rapporto tra varianze. Il suo uso non è però limitato al semplice modello di analisi esposto in queste pagine. Un modello più complesso di analisi della varianza utilizzato frequentemente in biologia è esposto alla fine del capitolo. Un altro, a supporto dell'analisi della regressione, sarà illustrato nel prossimo capitolo.



Esempio di calcolo:

	TOTALE	TRA GRUPPI	ENTRO GRUPPI
	$(x-m_{GEN})^2$	$(m_{GRUP}-m_{GEN})^2$	$(x-m_{GRUP})^2$
Gruppo A			
1	$(1-6)^2=25$	$(2-6)^2=16$	$(1-2)^2=1$
2	$(2-6)^2=16$	$(2-6)^2=16$	$(2-2)^2=0$
3	$(3-6)^2=9$	$(2-6)^2=16$	$(3-2)^2=1$
	S=50	S=48	S=2
$m_A=2$			
Gruppo B			
4	$(4-6)^2=4$	$(6-6)^2=0$	$(4-6)^2=4$
6	$(6-6)^2=0$	$(6-6)^2=0$	$(6-6)^2=0$
8	$(8-6)^2=4$	$(6-6)^2=0$	$(8-6)^2=4$
	S=8	S=0	S=8
$m_B=6$			
Gruppo C			
9	$(9-6)^2=9$	$(10-6)^2=16$	$(9-10)^2=1$
10	$(10-6)^2=16$	$(10-6)^2=16$	$(10-10)^2=0$
11	$(11-6)^2=25$	$(10-6)^2=16$	$(11-10)^2=1$
	S=50	S=48	S=2
$m_C=10$			
$\Sigma_{TOTALE}=54$			
$m_{TOTALE}=6$			
	TOTALE	TRA GRUPPI	ENTRO GRUPPI
DEVIANZA (S)	50+8+50=**108**	48+0+48=**96**	2+8+2=**12**
GDL	9-1=**8**	3-1=**2**	2+2+2=**6**
VARIANZA (s^2)	108/8=**13.5**	96/2=**48**	12/6=**2**

Come si può notare, la devianza totale è pari alla somma delle devianze tra ed entro gruppi:

108 = 96 + 12

come pure i rispettivi gradi di libertà:

8 = 2 + 6

Notare anche varianze tra ed entro gruppi non sono da sommare, ma da mettere in rapporto per avere appunto il test F:

$$F_{2,6\ GDL} = \frac{48}{2} = 24$$

Il valore di 24 va confrontato con quello riportato nella tabella della distribuzione F in corrispondenza di 2 gradi di libertà per il numeratore (= gradi di libertà tra gruppi) e 6 gradi di libertà per il denominatore (= gradi di libertà entro gruppi). 24 risulta superiore sia al il valore di F per α=0.05 è (F=5.14), sia a quello per α=0.01 (F=10.92). Si concluderà quindi che le tre medie dei tre gruppi in questione potrebbero provenire dalla stessa popolazione con una probabilità minore di uno su cento. Pertanto si rifiuta l'ipotesi nulla e si accetta l'ipotesi alternativa.

Valori critici di F corrispondenti a P≤0.05 per alcuni gradi di libertà											
	GDL del numeratore	1	2	3	4	5	...	10	...	50	...
GDL del denominatore											
1		161	200	216	225	230		242		252	
2		18.51	19.00	19.16	19.25	19.30		19.39		19.47	
3		10.13	9.55	9.28	9.12	9.01		8.78		8.58	
4		7.71	6.94	6.59	6.39	6.26		5.96		5.70	
5		6.61	5.79	5.41	5.19	5.05		4.74		4.44	
6		5.99	5.14	4.76	4.53	4.39		4.06		3.75	
...											
10		4.69	4.10	3.71	3.48	3.33		2.97		2.64	
...											
50		4.03	3.18	2.79	2.56	2.40		2.02		1.60	
...											

Valori critici di F corrispondenti a P≤0.01 per alcuni gradi di libertà											
	GDL del numeratore	1	2	3	4	5	...	10	...	50	...
GDL del denominatore											
1		4052	4999	5403	5625	5764		6056		6302	
2		98.49	99.00	99.17	99.25	99.30		99.40		99.48	
3		34.12	30.82	29.46	28.71	28.24		27.23		26.35	
4		21.20	18.00	16.69	15.98	15.52		14.54		13.69	
5		16.26	13.27	12.06	11.39	10.97		10.05		9.24	
6		13.47	10.92	9.78	9.15	8.75		7.87		7.09	
...											
10		10.04	7.56	6.55	5.99	5.64		4.85		4.12	
...											
50		7.17	5.06	4.20	3.72	3.41		2.70		1.94	
...											

Come si vede, la tabella di F è a tre entrate: i gradi di libertà del numeratore, i gradi di libertà del denominatore ed il livello di probabilità α. Ad esempio, il valore critico di F per un rapporto di varianze con 3 GDL al numeratore, 5 GDL al denominatore ed un livello probabilità α=0.05 è 5.41. Ciò si può indicare più semplicemente con

$F_{3, 5, .05} = 5.41$

Esercizio

		TOTALE	TRA GRUPPI	ENTRO GRUPPI
	x	$(x-m_{TOT})^2$	$(m_{GRUPPO}-m_{TOT})^2$	$(x-m_{GRUPPO})^2$
Gruppo A m_A =		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
Gruppo B m_B =		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
Gruppo C m_C =		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
		$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$	$(\quad - \quad)^2 =$
	m_{TOT} =	S_{TOT}=	S_{TRA}=	S_{ENTRO}=
		$GDL_{TOT} = n_{TOT}-1 =$	$GDL_{TRA} = n_{GRUPPI}-1 =$	$GDL_{ENTRO} = n_{TOT}-n_{GRUPPI} =$
		$s^2_{TOT} =$	$s^2_{TRA} =$	$s^2_{ENTRO} =$
	F =			

Il problema dei confronti multipli

Il risultato dell'analisi della varianza lascia spesso insoddisfatti. Infatti, anche se il valore di F avvalora l'ipotesi alternativa, il test non ci dice quale o quali gruppi differiscano dagli altri. Il test F è un test globale. E' come una spia che si accende quando succede qualcosa, ma non sappiamo esattamente cosa. Spesso desideriamo saperne di più, e fare tutti i possibili confronti tra tutti i gruppi presi a due a due. La tentazione sarebbe quella di fare diversi test t tra tutte le possibili coppie di medie, e molti cedono a tale tentazione. L'errore in questo caso sta nel fatto che un test t dà una differenza come significativa quando la probabilità a favore dell'ipotesi nulla scende sotto il valore di 0.05. Il rischio di sbagliarsi, rifiutando l'ipotesi nulla, è quindi del 5% o meno. Ma se, in uno stesso studio, ci si appoggia ai risultati di diversi test t effettuati confrontando un certo numero di medie, il rischio di falsi positivi aumenta in proporzione con il numero di confronti. Pertanto è scorretto l'uso del test *t* per saggiare simultaneamente tutte le differenze tra le possibili coppie di medie di un certo numero di gruppi. Questo problema, spesso non ben afferrato, passa sotto il nome di problema dei confronti multipli. Vediamo ora come può essere risolto.

La soluzione drastica di Bonferroni

Per rimediare all'aumento del rischio di falsi positivi la soluzione di Bonferroni è quella di riferirsi ad un valore soglia α inversamente proporzionale al numero di confronti multipli. Dati N gruppi, i possibili confronti tra tutte le medie prese a 2 a 2 sono $N(N-1)/2 = (N^2-N)/2$. Supponendo, ad es., un numero di 6 medie, i confronti saranno $6(6-1)/2 = 15$. Perciò bisognerebbe modificare la soglia di significatività da $\alpha=0.05$ ad $\alpha=0.05/15=0.0033$. Tale criterio è eccessivamente severo o **conservativo**, nel senso che conserva troppo l'ipotesi nulla, abbassando troppo il rischio (α) di falsi positivi ed elevando troppo quello (β) di falsi negativi. Il motivo sta essenzialmente nel fatto che se anche i confronti aumentano in ragione di N^2, le medie da confrontare sono sempre le stesse N medie, per cui i confronti, per quanto multipli, non sono del tutto indipendenti. Vi è infatti una relazione gerarchica tra i valori delle medie che il criterio di Bonferroni trascura. Per questo, nella pratica, quando è possibile si scelgono soluzioni alternative al criterio di Bonferroni, che comunque ha il merito storico e didattico di introdurre il problema dei confronti multipli.

Il test q di Student-Newman-Keuls (SNK)

Per eseguire il test di SNK occorre innanzitutto disporre le medie da confrontare in una graduatoria in ordine crescente. Quindi, per ogni confronto tra due medie di due gruppi **a** e **b**, bisogna valutare:

n_a e n_b : la numerosità dei due gruppi
s^2_{entro} : la varianza entro gruppi, calcolata preliminarmente per tutti i gruppi
p : il numero di medie comprese in graduatoria tra le due a confronto (includendo nel numero anche le due in esame; se queste sono immediatamente adiacenti p=2)

A questo punto è possibile calcolare la statistica q applicando la formula:

$$q = \frac{m_a - m_b}{\sqrt{\frac{s^2_{entro}}{2} \cdot \left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$$

Si entra in tabella con i gradi di libertà della varianza entro gruppi ed il parametro p. Come al solito, se il valore di q calcolato supera quello tabulato al livello critico di α=0.05, si rigetterà l'ipotesi nulla e si considererà significativa la differenza tra le due medie.

Valori critici di q corrispondenti a P≤0.05 per alcuni gradi di libertà e alcuni valori di p

	p	2	3	4	5	6	7	8	9	10	...
GDL s^2_{entro}											
1		17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	
2		6.09	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	
3		4.50	5.91	6.83	7.50	8.04	8.48	8.85	9.18	9.46	
4		3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	
5		3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	7.00	
6		3.46	4.34	5.90	5.31	5.63	5.90	6.12	6.32	6.49	
...											
10		3.15	3.88	4.33	4.65	4.91	5.12	5.31	5.46	5.60	
...											
60		2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	
...											

Valori critici di q corrispondenti a P≤0.01per alcuni gradi di libertà e alcuni valori di p

	p	2	3	4	5	6	7	8	9	10	...
GDL s^2_{entro}											
1		90.03	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6	
2		14.04	19.02	22.29	24.72	26.73	28.20	29.53	30.68	31.69	
3		8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69	
4		6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27	
5		5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	
6		5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	
...											
10		4.48	5.27	5.77	6.14	6.43	6.67	9.88	7.06	7.21	
...											
60		3.76	4.28	4.60	4.82	4.99	5.13	5.25	5.36	5.45	
...											

Un esempio di analisi della varianza applicata ad un disegno sperimentale complesso

All'inizio del capitolo abbiamo visto i principi fondamentali dell'analisi della varianza. In particolare abbiamo esaminato il modello più semplice di analisi della varianza, quello che viene comunemente detto *a una via (one way analysis of variance)*. Tuttavia l'analisi della varianza è un'analisi molto flessibile, applicabile a situazioni sperimentali anche molto complesse. In tal caso anche lo schema dell'esperimento (*disegno sperimentale*) è molto importante per ricavare il massimo dell'informazione statistica.
Un esempio di analisi della varianza un tantino complesso (per fornire un'idea) è il seguente:
Supponiamo di voler sperimentare l'effetto di due farmaci sulla pressione arteriosa.

- Otto animali ricevono i due farmaci (A e B) in due diverse dosi (1× e 2×). Queste variabili sono dette di controllo o fattori o...(vedi sinonimi sotto). Poiché diversi animali sono trattati in modo diverso, i confronti tra farmaci e dosi sono confronti TRA gruppi.
- I valori di pressione arteriosa (variabile dipendente) sono letti dopo 1, 2 e 3 ore dal trattamento. Questo consente di stabilire se l'effetto del trattamento varia in funzione del tempo. Poiché le misure vengono ripetute sugli stessi animali, la valutazione dell'effetto del tempo trascorso dal trattamento rappresenta un confronto ENTRO gruppi.
- Poiché si ritiene che gli animali possano essere più o meno sensibili ai due farmaci in relazione all'età ed al sesso, è possibile inserire queste variabili (covariate) nell'analisi al fine di escluderne l'interferenza sulla variabile dipendente e ridurre l'errore. Il mancato inserimento delle covariate può seriamente alterare il risultato dell'analisi (talvolta anche aumentando la probabilità di falsi positivi). E' invece assolutamente necessario che i fattori non influiscano sulle covariate o viceversa (nel nostro caso è assai improbabile che il trattamento influenzi l'età o il sesso degli animali o viceversa)
- Infine, è utile anche considerare alcuni animali di controllo. Si tratta di un gruppo al di fuori del contesto dello schema fattoriale. E' il cosiddetto gruppo isolato o appeso: *hanging group*, importante per verificare l'effetto dei farmaci rispetto alla condizione basale.

	VARIABILI (sinonimi) •DI CONTROLLO •DI RAGGRUPPAMENTO •INDIPENDENTI •FATTORI		VARIABILE (sinonimi) •DIPENDENTE •RISPOSTA CON MISURE RIPETUTE			COVARIATE	
animali	tipo di farmaco	dose del farmaco	valori di dopo 1 ora	pressione dopo 2 ore	arteriosa dopo 3 ore	età	sesso
1	A	1×	...	...	...	...	...
2	A	1×	...	...	...	...	...
3	A	2×	...	...	...	...	...
4	A	2×	...	...	...	...	...
5	B	1×	...	...	...	...	...
6	B	1×	...	...	...	...	...
7	B	2×	...	...	...	...	...
8	B	2×	...	...	...	...	...
9	controllo	controllo	...	...	...	...	...
10	controllo	controllo	...	...	...	...	...
11	controllo	controllo	...	...	...	...	...
12	controllo	controllo	...	...	...	...	...

L'analisi della varianza applicata a tale esperimento consente la valutazione

degli **effetti principali**
- tipo di farmaco *(i due farmaci hanno effetti diversi sulla pressione arteriosa ?)*
- dose *(le dosi utilizzate danno effetti diversi ?)*
- tempo *(l'effetto varia col tempo ?)*

delle **interazioni** tra variabili
- tipo di farmaco × dose *(un certo tipo di farmaco ha un effetto particolare se somministrato ad una certa dose ?)*
- tipo di farmaco × tempo *(un certo tipo farmaco ha un effetto particolare dopo un certo lasso di tempo dalla somministrazione ?)*
- dose del farmaco × tempo *(una certa dose ha un effetto particolare dopo un certo lasso di tempo dal trattamento ?)*
- farmaco × dose × tempo *(un certo tipo di farmaco ha un effetto particolare dopo un certo lasso di tempo dal trattamento e se somministrato in una certa dose ?)*

delle **interazioni** tra variabili e covariate
- tipo di farmaco × età *(un certo tipo di farmaco ha un effetto particolare su soggetti di diversa età ?)*
- dose del farmaco × sesso *(un certo tipo di farmaco ha un effetto particolare su soggetti di un certo sesso ?)*
- ecc. ecc.

di **specifici confronti** (o contrasti) tra diversi gruppi di animali trattati ed il gruppo di animali di controllo.

ULTIMI AVVISI AI NOVIGANTI

L'analisi della varianza è spesso detta ANOVA (*an*alysis *o*f *va*riance)
Quando esistono due o più variabili di controllo si parla di ANOVA a due o più vie.
Quando si utilizzano covariate si può parlare (non è obbligatorio) di ANCOVA.
MANOVA (*m*ultivariate *an*alysis *o*f *va*riance) è invece un'ANOVA con più variabili dipendenti e MANCOVA quella con più variabili dipendenti e covariate.