

Lezione 13

L'analisi della Varianza (ANOVA): il modello lineare

Argomenti della lezione:

- ➔ **Modello lineare**
- ➔ **Disegni a una via**

L'Analisi della Varianza (ANOVA):

Esamina differenze tra le medie di due o più popolazioni

Variabile Dipendente ➔ Intervalli Equivalenti

Variabile Indipendente ➔ Catoriale

Diversi modelli di ANOVA:

- ➔ **Una sola V.I.: Disegni a una via**
- ➔ **Due o più V.I.: Disegni Fattoriali**
- ➔ **Una sola V.D.: Analisi univariata**
- ➔ **Due o più V.D.: Analisi multivariata (MANOVA)**

L'analisi della varianza univariata (ANOVA):

disegni tra i soggetti ad un solo fattore

Disegni "tra i soggetti" o "per gruppi indipendenti": ad ogni trattamento o condizione sperimentale (livelli della variabile indipendente) corrisponde un diverso gruppo di soggetti

In ogni gruppo (condizione) ci sono soggetti diversi

OBBIETTIVI	
SI	NO
S1	S6
S2	S7
...	...

Modello lineare dell'ANOVA

Punteggio y_{ij} di un soggetto "j" nel gruppo "i" :

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ dove:}$$

- μ : media generale
- α_i : effetto del trattamento
- ε_{ij} : componente "residua"

Stime campionarie dei parametri di popolazione

$$\hat{\mu} = \bar{y}_{..}$$

$$\hat{\alpha}_i = \bar{y}_{i.} - \bar{y}_{..}$$

$$\hat{\varepsilon}_{ij} = y_{ij} - \bar{y}_{i.}$$

Il punteggio del soggetto ij è esprimibile in questo modo:

$$y_{ij} = \bar{y}_{..} + (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

$$\hat{\mu} + \hat{\alpha}_i + \hat{\varepsilon}_{ij}$$

Scarto dalla media:

$$(y_{ij} - \bar{y}_{..}) = (\bar{y}_{i.} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{i.})$$

$$\text{Devianza Totale} = SS_T = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2$$

$$\text{Devianza Tra i Gruppi} = SS_B = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$\text{Devianza Entro i Gruppi} = SS_W = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

**SS = Sum of Squares,
Somma dei quadrati**

È possibile dimostrare che la scomposizione vista per gli scarti dalla media vale anche per le devianze, ovvero:

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$$

ovvero:

$$SS_T = SS_B + SS_W$$

Per ognuna delle tre devianze è possibile definire un opportuno numero di gradi di libertà:

SS_T : n-1 gdl

SS_B : k-1 gdl

SS_W : n-k gdl

$$\text{gdl}(SS_T) = \text{gdl}(SS_B) + \text{gdl}(SS_W)$$

$$n-1 = (k-1) + (n-k)$$

Devianza diviso gradi di libertà = varianza (Mean Square)

Varianza totale $MS_T = SS_T / (n-1)$

Varianza tra i gruppi $MS_B = SS_B / (k-1)$

Varianza entro i gruppi $MS_W = SS_W / (n-k)$

Per le varianze non vale la scomposizione che vale per le devianze e per i gradi di libertà:

$$MS_T \neq MS_B + MS_W$$

ANOVA effettuata a mano: formule "abbreviate" (Ercolani, Areni e Leone, 2002)

Rapporto "F"

Il rapporto tra le varianze MS_B / MS_W segue la distribuzione F di Snedecor/Fisher

Ipotesi statistiche:

H_0 : Le medie sono uguali

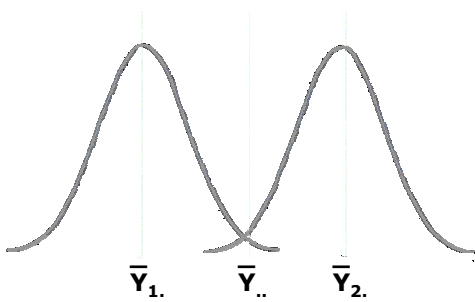
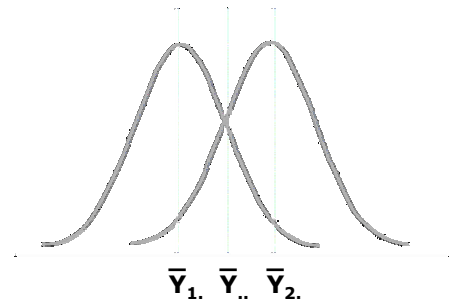
H_1 : Almeno due medie diverse

La varianza tra i gruppi riflette l'effetto del trattamento

La varianza entro i gruppi, o within, non può essere spiegata in base agli effetti del trattamento, viene attribuita all'errore casuale

H_0 vera \Rightarrow varianze tra i gruppi ed entro i gruppi molto simili, rapporto vicino ad 1. MS_B e MS_W sono due stime parallele e indipendenti della varianza totale della popolazione (MS_T)

H_0 falsa \Rightarrow varianza tra i gruppi maggiore della varianza entro i gruppi. Rapporto F elevato ($MS_B > MS_W$)

F significativa (Rifiuto H_0)**F non significativa (Accetto H_0)****Assunzioni**

- ↻ gli errori (ε_{ij}) devono seguire la distribuzione normale ed avere media uguale a 0
- ↻ la varianza degli errori (σ_e) deve essere uguale in ogni gruppo (condizione di omoschedasticità)

Assunzioni

- ↻ gli errori (ε_{ij}) devono essere indipendenti (il punteggio di un soggetto non deve essere correlato con quello di altri soggetti)
- ↻ gli effetti hanno una natura additiva

Esempio

Verificare l'efficacia di programmi di formazione che prevedono:

- 📖 Obiettivi (condizione A)
- 📖 Obiettivi + feedback (condizione B)
- 📖 Condizione di controllo (condizione C)

VD = numero di problemi risolti

Obiettivi (Y_1): 10 7 4 5 8

Obiettivi + Feedback (Y_2): 9 10 5 4 7

Controllo (Y_3): 3 2 2 3 1

Ipotesi statistiche

(H₀): $\mu_1 = \mu_2 = \mu_3$
 (le 3 medie sono relative a campioni che provengono dalla stessa popolazione)

(H₁): almeno due μ diverse,
 $\mu_1 \neq \mu_2$, $\mu_1 \neq \mu_3$, $\mu_2 \neq \mu_3$
 (almeno due medie sono relative a campioni che provengono da popolazioni diverse)

Valore critico di F

Gradi di libertà al numeratore, della varianza tra i gruppi:

$$(k - 1) = 3 - 1 = 2$$

Gradi di libertà al denominatore, della varianza entro i gruppi:

$$(n - k) = 15 - 3 = 12$$

$$F(2, 12) = 3.89, \alpha = 0.05$$

$$F(2, 12) = 6.93, \alpha = 0.01$$

Calcolo delle varianze e del rapporto F

G _i	(Y _{ij} - $\bar{Y}_{..}$)	($\bar{Y}_{i.}$ - $\bar{Y}_{..}$)	(Y _{ij} - $\bar{Y}_{i.}$)
10	(10-5.3)	(5.3-5.3)	(10-5.3)
7	(7-5.3)	(5.3-5.3)	(7-5.3)
4	(4-5.3)	(5.3-5.3)	(4-5.3)
5	(5-5.3)	(5.3-5.3)	(5-5.3)
3	(3-5.3)	(5.3-5.3)	(3-5.3)

$$SS_B = (6.8-5.3)^2 + \dots + (7-5.3)^2 + \dots + (2.2-5.3)^2 = 73.73$$

$$MS_B = 73.73 / 2 = 36.87$$

$$SS_W = (10-6.8)^2 + \dots + (9-7)^2 + \dots + (3-2.2)^2 = 51.6$$

$$MS_W = 51.6 / 12 = 4.3$$

$$F(2,12) = 36.87 / 4.3 = 8.57$$

$$F_{critico}(2, 12) = 6.93, \alpha = 0.01$$

Rifiuto H₀

Tabella riassuntiva dei risultati dell'ANOVA

Fonte	SS	GDL	MS	F	p
Between	73.73	2	36.87	8.57	.01
Errore	51.60	12	4.30		
Totale	125.33	14			

Confronti tra le medie dei gruppi

CONDIZIONE	Media	DS.	N
A (Obiettivi)	6.8	2.4	5
B (Ob.+ Feed.)	7.0	2.5	5
C (Controllo)	2.2	.84	5
Totale	5.3	2.99	15

I confronti post hoc

Ogni media viene confrontata con tutte le altre. Il ricercatore non seleziona in anticipo, tra tutti i confronti possibili, quelli rilevanti ai fini della sua ipotesi

Aumento della probabilità di commettere l'errore di primo tipo

Condizione A vs. Condizione B, A vs. C, B vs. C.

Ipotesi: almeno un confronto risulti significativo

$\alpha = 0.05$ per ogni confronto

Probabilità che almeno uno dei tre confronti risulti significativo = $.05 + .05 + .05 = .15$

Livello reale di $\alpha = 3 * .05 = .15$.

In generale, in un set di k confronti post-hoc esaminati simultaneamente, il livello reale di probabilità che almeno uno di essi risulti significativo non è α ma $k\alpha$

Soluzione: scegliere un valore α minore di $.05$ (es., $.05/3 = .017$, e in genere $.05/k$)

Procedura post-hoc Tukey HSD

(Per le formule vedi "Analisi dei Dati", p. 202)

→ Controllo vs. Obiettivi, diff=-4.6**

→ Controllo vs. Ob.+Feed., diff=-4.8**

→ Obiettivi vs. Ob.+Feed., diff=-.2 (ns)

** $p < .01$

I confronti pianificati

Effettuare solo i confronti che appaiono più rilevanti ai fini dell'ipotesi di ricerca

Differenza tra 2 medie:

- singoli gruppi
- media "aggregata" ottenuta combinando più gruppi

Primo confronto:

"controllo" vs. "obiettivi" e "obiettivi + feedback" combinati insieme

Secondo confronto:

"obiettivi" vs. "obiettivi + feedback"

Per effettuare i confronti si deve attribuire ad ogni media un coefficiente positivo o negativo. Le somme dei coeff. devono essere uguali a 0

Medie con segno diverso \Rightarrow confrontate tra loro

Medie con segno uguale \Rightarrow combinate assieme

Medie con coefficiente 0 \Rightarrow escluse dal confronto

Coefficienti per i 2 confronti

	Ob.	Ob. + Feed.	Contr.
1° confronto	-1	-1	2
2° confronto	1	-1	0

Risultati dei due confronti

	SS	gdl	MS	F
1° confronto				
Confronto	73.63	1	73.63	17.12**
Errore	51.60	12	4.30	
2° confronto				
Confronto	.10	1	.10	.23 (ns)
Errore	51.60	12	4.30	

Confronti Ortogonali

\rightarrow Forniscono informazioni indipendenti: il risultato del primo non dà indicazioni sul risultato del secondo, e viceversa

\rightarrow Il numero massimo di confronti ortogonali è uguale a $k - 1$

Confronti Ortogonali

\rightarrow La somma dei prodotti dei coefficienti dei confronti deve essere uguale a 0

\rightarrow Anche le somme dei coefficienti devono essere uguali a 0

Confronti Ortogonali

	Ob.	Ob. + Feed.	Contr.	Somme
1° confronto	-1	-1	2	0
2° confronto	1	-1	0	0
Prodotti	-1	1	0	0

CONCLUSIONE

- **Modello lineare**
- **Disegni ad una via tra i soggetti**
- **Confronti post-hoc e pianificati**