

Analisi della Varianza

(ad un criterio)

ANOVA

(Analysis Of Variance)

1

ANALISI DELLA VARIANZA - 1

Abbiamo k gruppi, con un numero variabile di unità statistiche. Nella notazione classica, ogni unità statistica viene individuata da due numeri in posizione pedice: il primo indica il gruppo di appartenenza, e il secondo indica la posizione del soggetto all'interno del gruppo.

gruppo 1	gruppo 2	gruppo 3	gruppo k
X_{11}	X_{21}	X_{31}	X_{k1}
X_{12}	X_{22}	X_{32}	X_{k2}
X_{13}	X_{23}	X_{33}	X_{k3}
X_{14}	X_{24}	X_{34}	X_{k4}
X_{15}	X_{25}	X_{35}	X_{k5}
X_{16}	X_{26}	X_{36}	X_{k6}
X_{17}	X_{27}	X_{37}	X_{k7}
X_{18}		X_{38}	X_{k8}
		X_{39}	X_{k9}
\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_k

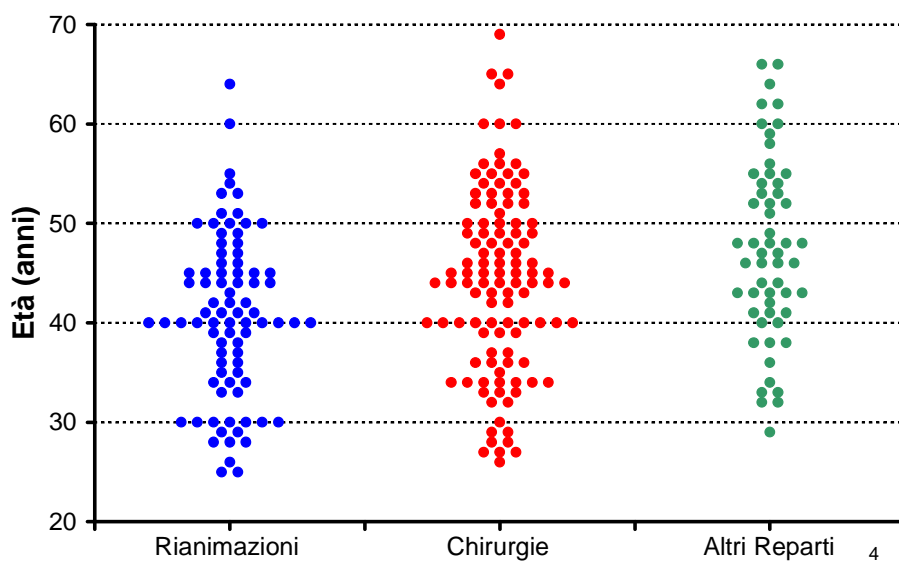
2

ANALISI DELLA VARIANZA - 2

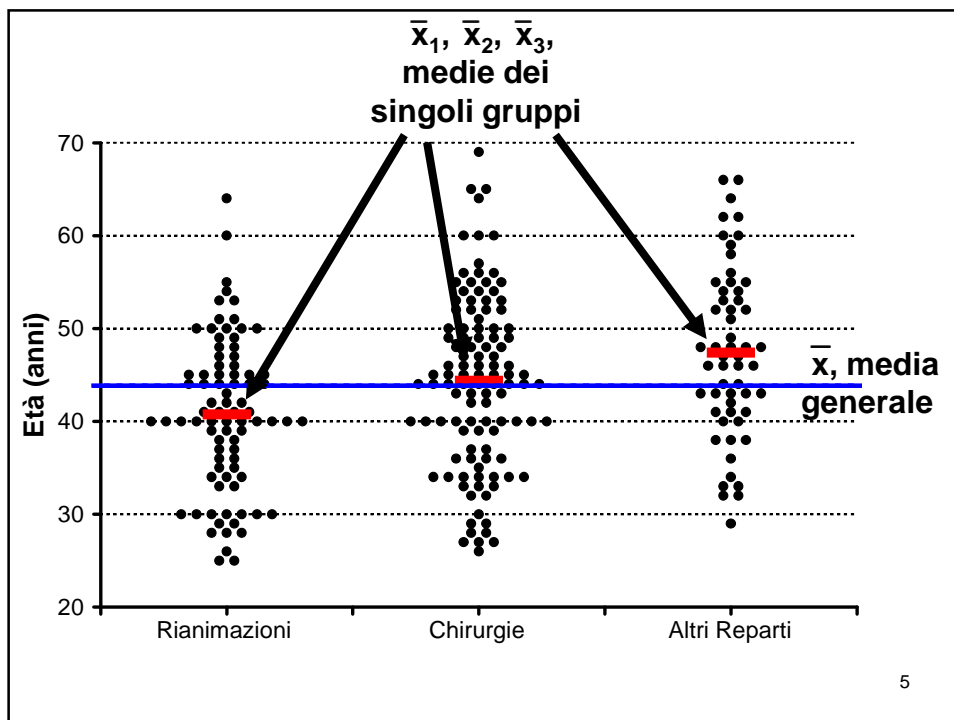
Oltre ad una media generale, \bar{x} , abbiamo k medie, una per ognuno dei singoli gruppi, $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_k$.

3

Età dei medici operanti nelle Rianimazioni, Chirurgie ed altri Reparti di un'Azienda Ospedaliera del Veneto



4



ANALISI DELLA VARIANZA - 3

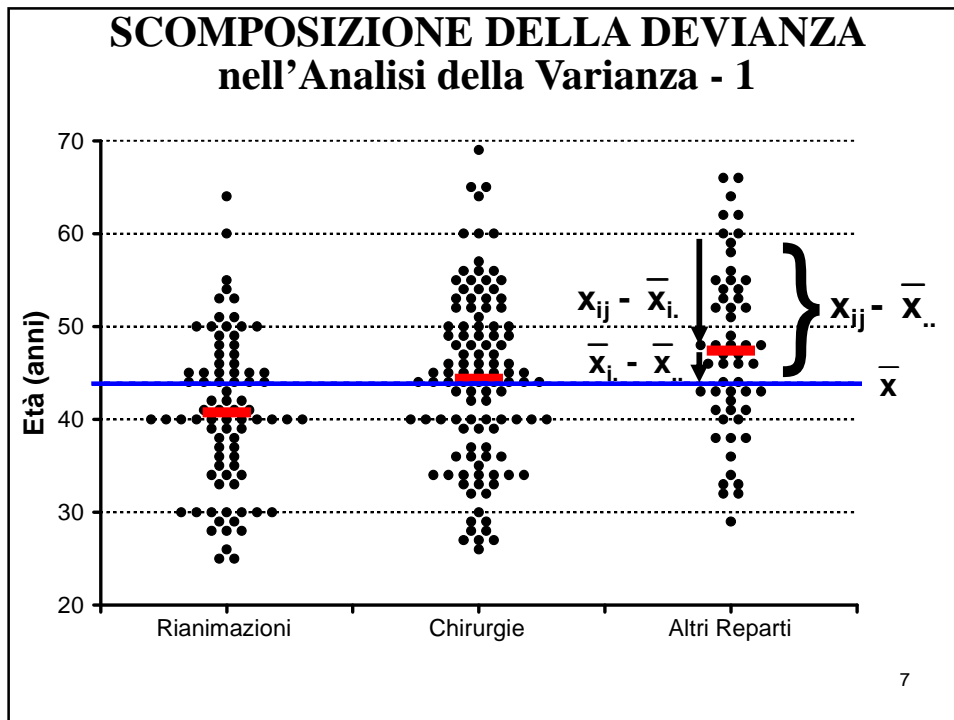
Ipotesi $\left\{ \begin{array}{l} H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_0 \\ H_1: \text{almeno una media differisce dalle altre} \end{array} \right.$

Per rispondere a questa domanda, possiamo fare tante t di Student, confrontando tutte le possibili coppie di medie?

NO, perché altrimenti avremmo un'inflazione (aumento anormale) di α (alfa), probabilità di errore del I tipo.

E' meglio quindi ricorrere ad un test globale, che confronti fra di loro tutti i gruppi:

l'analisi della varianza.



$x_{ij} - \bar{x}_{..}$ = scarto di una singola osservazione (valore *jesimo* del gruppo *iesimo*) dalla media generale

$\bar{x}_i - \bar{x}_{..}$ = scarto della media del gruppo *iesimo* dalla media generale

$x_{ij} - \bar{x}_i$ = scarto di una singola osservazione (valore *jesimo* del gruppo *iesimo*) dalla media del gruppo *iesimo*

SCOMPOSIZIONE DELLA DEVIANZA nell'Analisi della Varianza - 2

Per una singola osservazione:

$$\begin{array}{ccc}
 \text{Variabilità} & & \text{Variabilità} \\
 \text{totale} & & \text{entro gruppi} \\
 (x_{ij} - \bar{x}_{..}) & = & (\bar{x}_{i.} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{i.}) \\
 & & \text{Variabilità fra} \\
 & & \text{gruppi}
 \end{array}$$

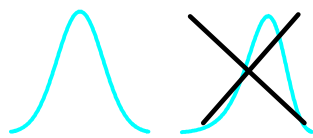
Si può dimostrare che, per tutte le osservazioni:

$$\begin{array}{ccc}
 \text{Devianza} & & \text{Devianza} \\
 \text{totale, SST} & & \text{entro gruppi} \\
 \sum_{i,j} (x_{ij} - \bar{x}_{..})^2 & = & \sum_{i,j} (\bar{x}_{i.} - \bar{x}_{..})^2 + \sum_{i,j} (x_{ij} - \bar{x}_{i.})^2 \\
 & & \text{Devianza tra} \\
 & & \text{gruppi}
 \end{array}$$

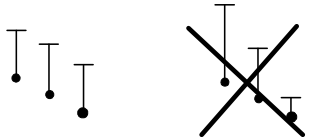
9

ASSUNZIONI PER TEST PARAMETRICI

NORMALITA'



OMOSCHEDASTICITA' (STABILITA' della VARIANZA)



INDEPENDENZA delle OSSERVAZIONI (ERRORI)

N = numero animali, non
numero di prove

t di Student
analisi della varianza
analisi della covarianza
regressione

10

Tabella dell'ANALISI DELLA VARIANZA
a 1 criterio di classificazione

Fonte di variabilità	Gradi di libertà	Devianza	Varianza	Test-F
TRA gruppi	k-1	$\sum n_i (\bar{x}_i - \bar{x}_{..})^2$	$DEV_{tra} / (k-1)$	$\frac{VAR_{tra}}{VAR_{entro}}$
ENTRO gruppi (errore)	N-k	$\sum \sum (x_{ij} - \bar{x}_i)^2$	$DEV_{entro} / (N-k)$	
Totale	N-1	$\sum \sum (x_{ij} - \bar{x}_{..})^2$		

Se H_1 e' vera (almeno 1 media e' diversa dalle altre), allora

$$\sigma_{tra}^2 > \sigma_{entro}^2 \approx \sigma^2$$

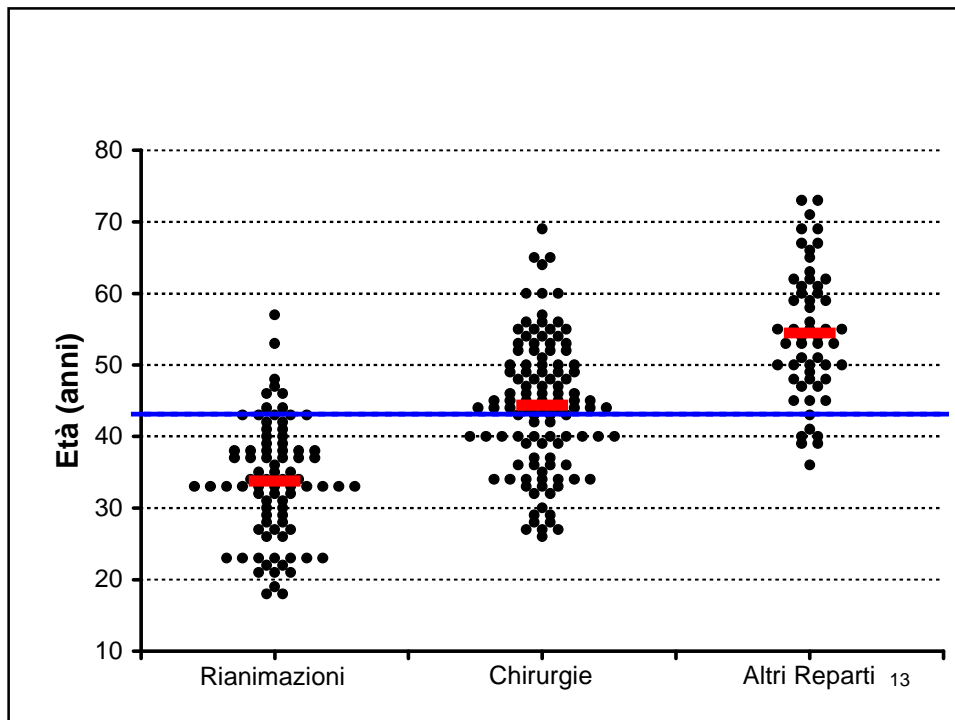
11

L'età dei medici (n=251) è significativamente diversa nelle Rianimazioni, nelle Chirurgie e negli altri Reparti?

fonte di variabilità	gradi di libertà	devianza	varianza	test F (significatività)
TRA gruppi	2	1546,10	773,05	9,779
ENTRO gruppi	248	19604,73	79,05	(P<0,001)
TOTALE	250	21150,84		

Si rifiuta l'ipotesi nulla: l'età differisce significativamente tra i vari Reparti.

12



Se il test F (il test globale) risulta significativo, si possono confrontare fra loro le singole medie, utilizzando dei test appropriati che vanno sotto il nome di confronti o contrasti multipli.

Questo tipo di analisi viene definita anche “post hoc”.

1) I contrasti multipli in genere utilizzano come stima della variabilità casuale la varianza residua, calcolata nell’ambito dell’analisi della varianza.

2) I contrasti multipli sono costruiti in modo tale da evitare l’inflazione di α , probabilità di errore del I tipo.

CONTRASTI (o CONFRONTI) MULTIPLI (POST HOC ANALYSIS)

TEST di SCHEFFE' e' il test piu' conservativo, si confrontano sia medie che gruppi di medie

TEST di TUKEY si confrontano tutte le possibili coppie di medie

TEST di DUNNETT si confrontano tutte le singole medie con un controllo

CORREZIONE di BONFERRONI

Si moltiplica il valore di 'p' (probabilita' di errore del I tipo) per il numero di test effettuati

Ad esempio, se effettuo 3 test statistici: p iniziale p corretta

0.03	*3=	0.09
0.15	*3=	0.45
0.01	*3=	0.03

15

- Se le assunzioni per effettuare l'analisi della varianza non sono verificate:
- Adeguata trasformazione della variabili oppure
- Test non parametrico di Kruskal-Wallis

16