

Analisi della Varianza

(ad un criterio)

ANOVA

(Analysis Of Variance)

ANALISI DELLA VARIANZA - 1

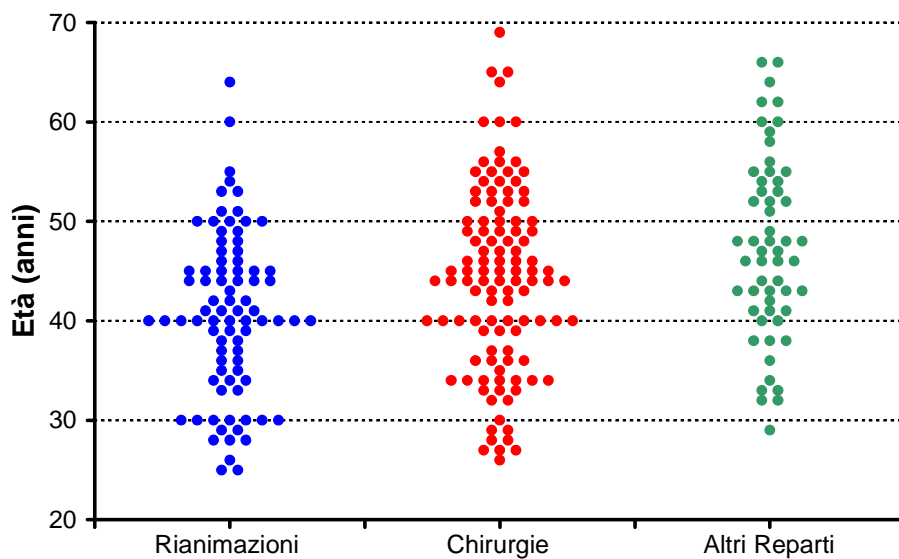
Abbiamo k gruppi, con un numero variabile di unità statistiche. Nella notazione classica, ogni unità statistica viene individuata da due numeri in posizione pedice: il primo indica il gruppo di appartenenza, e il secondo indica la posizione del soggetto all'interno del gruppo.

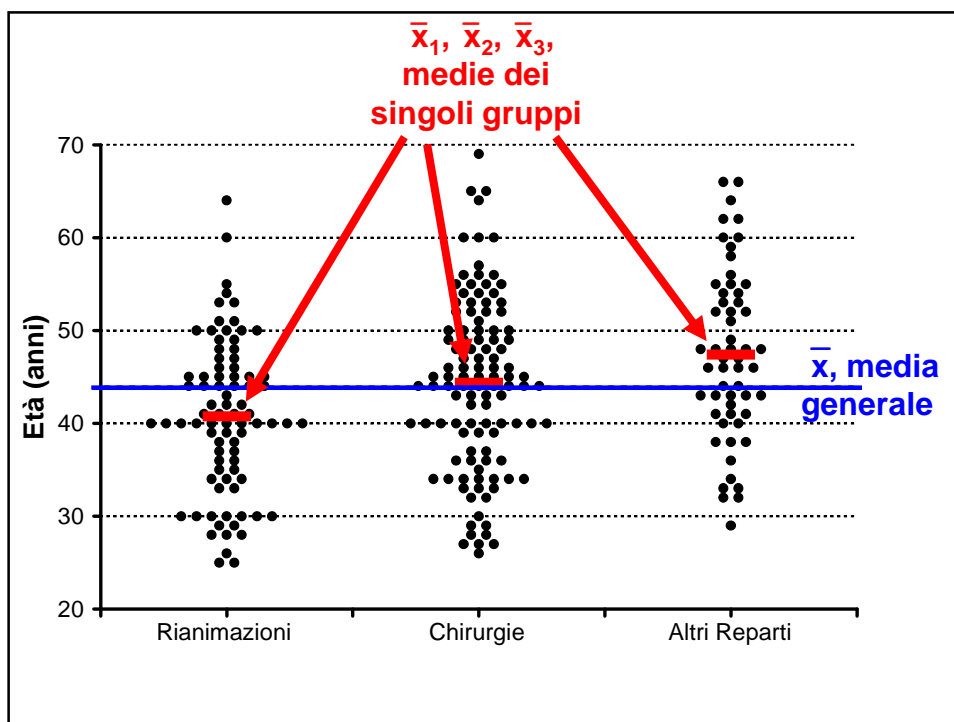
gruppo 1	gruppo 2	gruppo 3	gruppo k
X ₁₁	X ₂₁	X ₃₁	X _{k1}
X ₁₂	X ₂₂	X ₃₂	X _{k2}
X ₁₃	X ₂₃	X ₃₃	X _{k3}
X ₁₄	X ₂₄	X ₃₄	X _{k4}
X ₁₅	X ₂₅	X ₃₅	X _{k5}
X ₁₆	X ₂₆	X ₃₆	X _{k6}
X ₁₇	X ₂₇	X ₃₇	X _{k7}
X ₁₈		X ₃₈	X _{k8}
		X ₃₉	X _{k9}
\bar{X}_1	\bar{X}_2	\bar{X}_3	\bar{X}_k

ANALISI DELLA VARIANZA - 2

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

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





ANALISI DELLA VARIANZA - 3

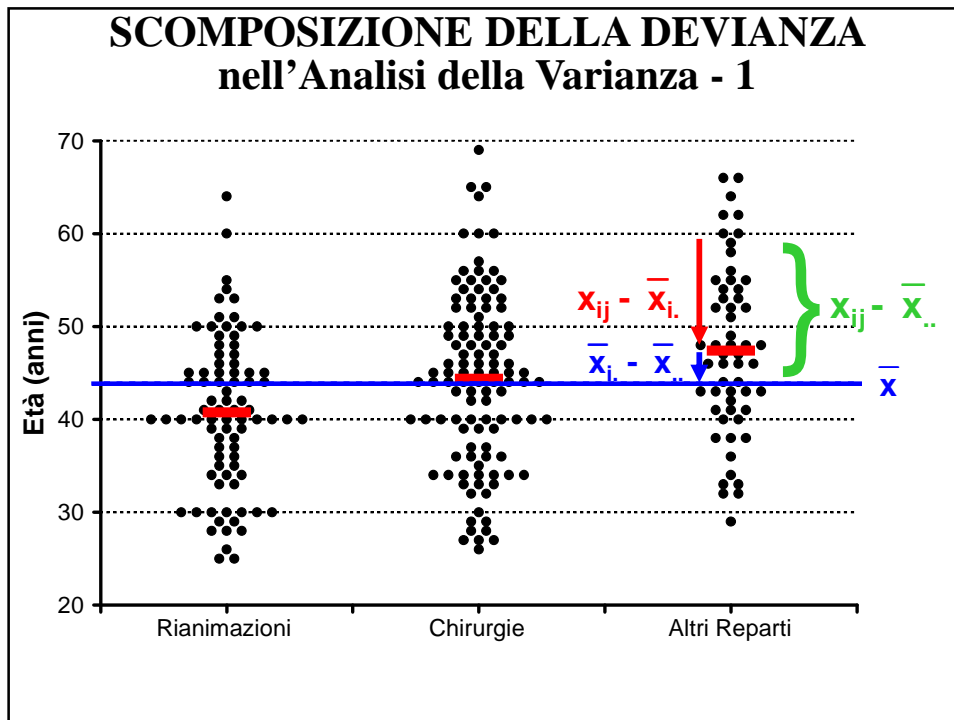
Ipotesi $\left\{ \begin{array}{l} H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_0 \\ H_1: \text{almeno una media differisce dalle altre} \end{array} \right.$

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 abnorme) 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:

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

Si può dimostrare che, per tutte le osservazioni:

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

Tavola ANOVA ad un criterio di classificazione

Fonte di variabilità	SS	df	MS
Tra gruppi	$\sum_{ij} (\bar{x}_{i.} - \bar{x}_{..})^2$	$k - 1$	$SS_{tra} / (k - 1)$
Entro gruppo	$\sum_{ij} (x_{ij} - \bar{x}_{i.})^2$	$N - k$	$SS_{entro} / (N - k)$
Totale	$\sum_{ij} (x_{ij} - \bar{x}_{..})^2$	$N - 1$	$SS_{tot} / (N - 1)$

VARIANZA DELLE MEDIE

VARIANZA POOLED

Scomposizione della devianza campionaria per fonti di variabilità:

- Differenze tra gruppi: SS_{tra}
- Errore campionario: SS_{entro}

I° STEP: definire l'ipotesi da verificare

$$\begin{cases} H_0: \mu_0 = \mu_1 = \dots = \mu_k \\ H_1: \mu_i \neq \mu_j \end{cases}$$

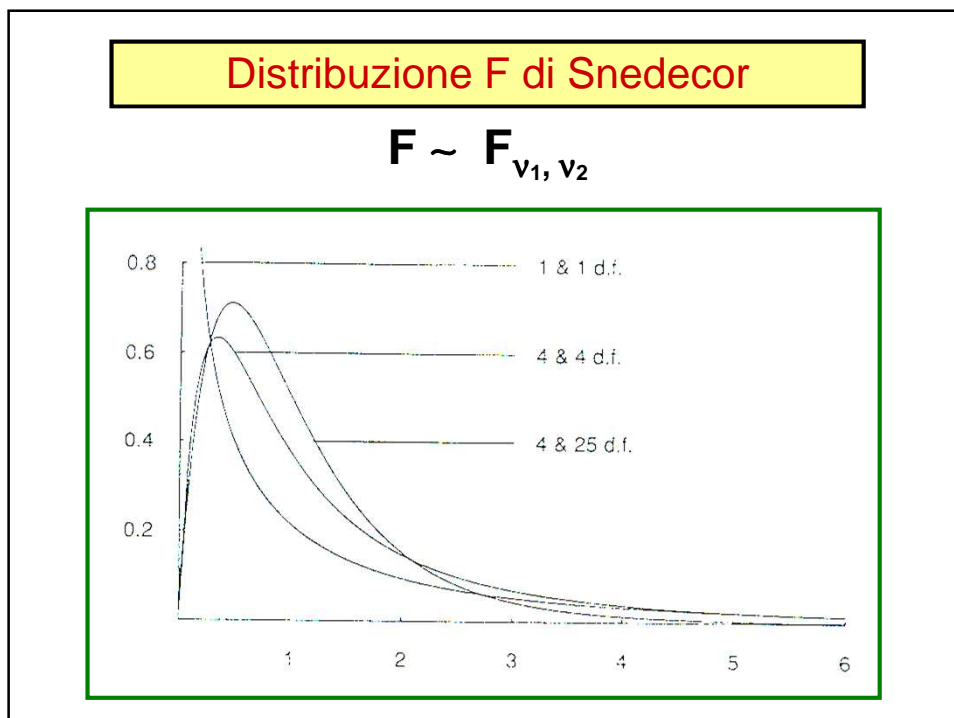
almeno una media
differente dalle altre

II° STEP: definire la statistica test → **III° STEP: calcolare il p-value**

$$F = \frac{MS_{tra}}{MS_{entro}} \underset{H_0}{\sim} F_{(k-1), (N-k)}$$

distribuzione F di Snedecor
con (k-1) gradi di libertà
al numeratore e (N-k) gradi di
libertà al denominatore

Se valgono le assunzioni di normalità e
omoschedasticità

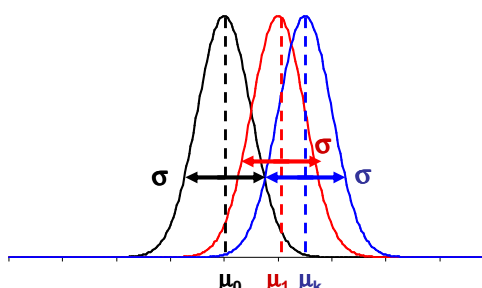


ASSUNZIONI SU CUI SI BASA L'ANALISI DELLA VARIANZA:
indipendenza

L'esito della misurazione in un soggetto non condiziona l'esito della misurazione negli altri soggetti.

omoschedasticità

$$\sigma_0^2 = \sigma_1^2 = \dots = \sigma_k^2 = \sigma^2$$

normalità


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.