



# Corso di Laurea di Scienze Forestali ed Ambientali

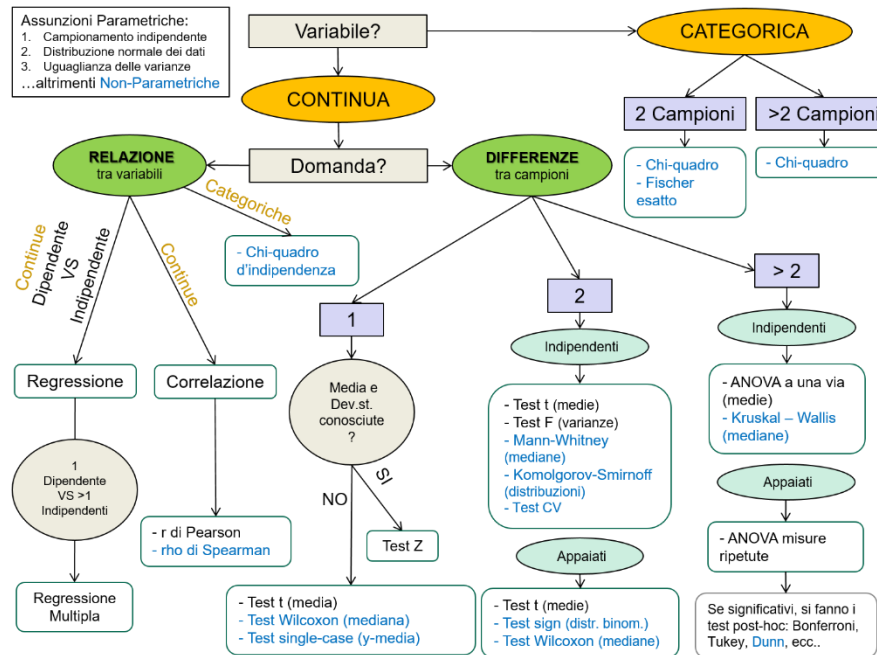
# Ecologia e Statistica

# per l'ambiente

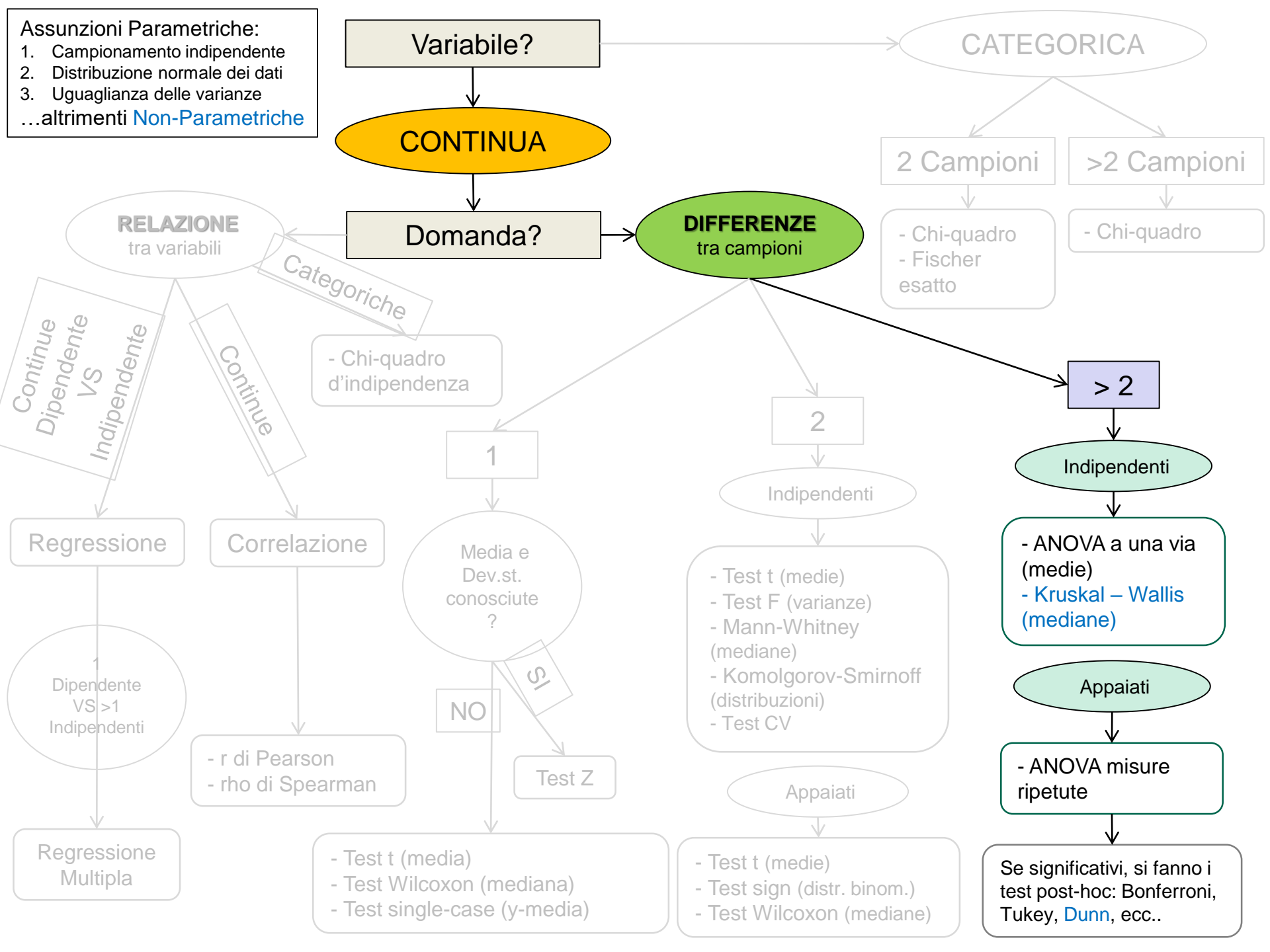


a.a. 2020-2021 - MATTEO GARBARINO - matteo.garbarino@unito.it

## STATISTICA INFERENZIALE



- Assunzioni Parametriche:**
1. Campionamento indipendente
  2. Distribuzione normale dei dati
  3. Uguaglianza delle varianze
- ...altrimenti **Non-Parametriche**



# Test per molti campioni

## *Several-sample tests*

L'analisi della varianza (ANalysis Of VAriance) o **ANOVA** è sicuramente il test statistico per più di 2 campioni più utilizzato in ecologia e non solo (es. agronomia).

# ANOVA

Sottoponiamo una coltura di frumento a **4 trattamenti** sperimentali differenti e misuriamone la produzione granellare.

Per meglio definire le popolazioni decidiamo di fare **5 ripetizioni** per ogni trattamento.

I 4 trattamenti vengono distribuiti nello spazio in modo randomizzato perché l'insieme degli effetti non dovuti al trattamento sia uguale in tutti i trattamenti.

2	3	3	1
1	4	4	2
1	2	2	2
3	1	3	1
4	3	4	4

# ANOVA

Abbiamo ottenuto i seguenti valori in t ha<sup>-1</sup>

5.8	4.7	5.6	4.4
6.5	5.9	9.4	5.3
6.8	4.1	5.5	5.8
7.1	7.4	6.5	6.4
6.2	6.1	4.5	7.6

Essi vengono ordinati nella tabella seguente:

Trat 1	Trat 2	Trat 3	Trat 4	Media generale
6.4	5.8	6.1	5.9	
4.4	5.3	7.1	6.2	
6.8	5.5	4.7	7.6	
6.5	5.8	6.5	4.5	
7.4	4.1	5.6	9.4	

Medie	6.3	5.3	6.0	6.7	6.075
-------	-----	-----	-----	-----	-------

I trattamenti hanno avuto effetto oppure no?

# ANOVA

Perché non usare il test t?

Nell'ipotesi di confrontare i 4 trattamenti 2 a 2 con il test t di Student, generando così 6 confronti, visto che ogni confronto ha una protezione pari a  $1 - \alpha$  (95% con  $\alpha = 5\%$ ), la protezione complessiva delle conclusioni combinate diventerebbe  $0.95 \times 0.95 \times 0.95 \times 0.95 \times 0.95 \times 0.95 = 0.735$

Quindi l'  $\alpha$  complessivo (probabilità totale di commettere un errore di tipo I) sarebbe  $1 - 0.735 = 0.265$ , cioè circa 27%!!!

Più in generale se  $n$  è il numero dei trattamenti, il numero di confronti possibili è:

$$(n*(n-1))/2$$

mentre la significatività reale complessiva di tutti i confronti è:

$$\alpha_T = 1 - ((1 - \alpha)^{(n*(n-1))/2})$$

La protezione non è quindi quella prevista. Occorre trovare un'altra via

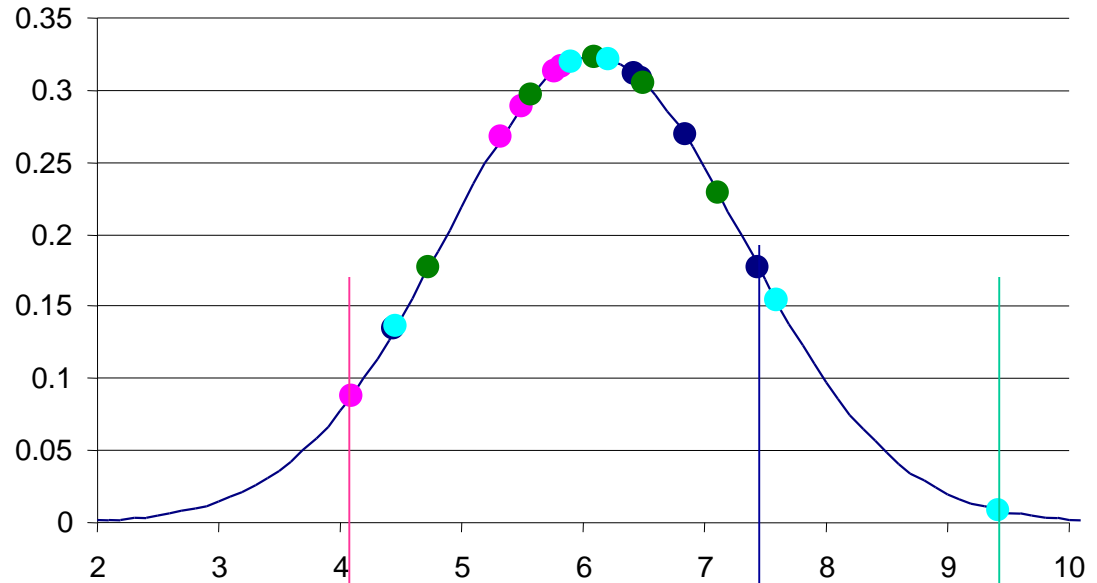
# ANOVA

Man mano che si aumenta il numero di confronti a coppie, aumenta la probabilità di un errore di Tipo I o  $\alpha$  (rigettare un'ipotesi vera)

Numero di Trattamenti	Numero di confronti a coppie	Probabilità Errore Tipo I ( $\alpha$ )
2	1	.05
4	6	.27
6	21	.66
10	45	.90

# Ipotesi nulla

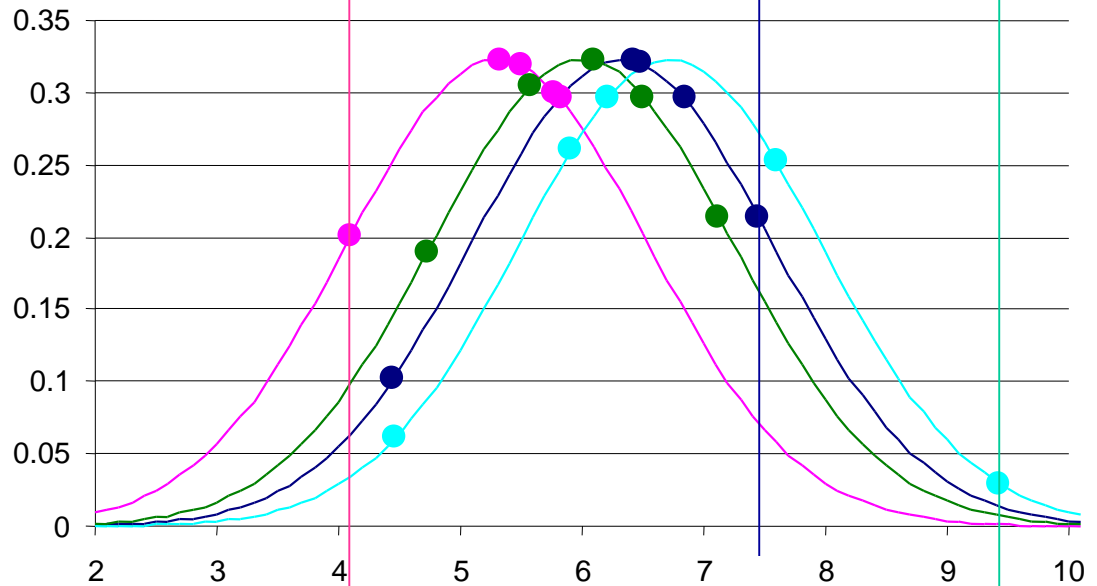
I valori medi dei 4 trattamenti differiscono tra loro solo per variazioni casuali

$$\mu_1 = \mu_2 = \mu_3 = \dots = \mu_n$$


# Ipotesi alternative

Esistono una o più differenze tra le medie delle popolazioni

Ad esempio:

$$\mu_1 \neq \mu_2 = \mu_3 = \dots = \mu_n$$




# ANOVA

ANOVA Permette di verificare la significatività delle differenze tra le medie aritmetiche di diversi gruppi di dati (campioni o trattamenti).

- **Trattamento:** fattore sperimentale, fonte o causa della variabilità dei dati
- **Ripetizione:** osservazioni che riflettono la variabilità dovuta ad altre variabili

Per poterla utilizzare devo avere almeno 2 ripetizioni per trattamento.

È una estensione del Test-t di Student e prevede un campionamento completamente randomizzato e si basa sulla **Distribuzione F**.

$$F = \frac{S^2 TRA}{S^2 IN}$$

Varianza TRA trattamenti (BETWEEN)

Varianza INTRA trattamenti (WITHIN)

# ANOVA: Varianza IN

Una prima varianza della popolazione (Varianza in) può essere calcolata come una “pooled variance” usando gli scarti di ogni individuo dalla sua media campionaria nell’ipotesi che le varianze siano omogenee fra tutti i trattamenti.

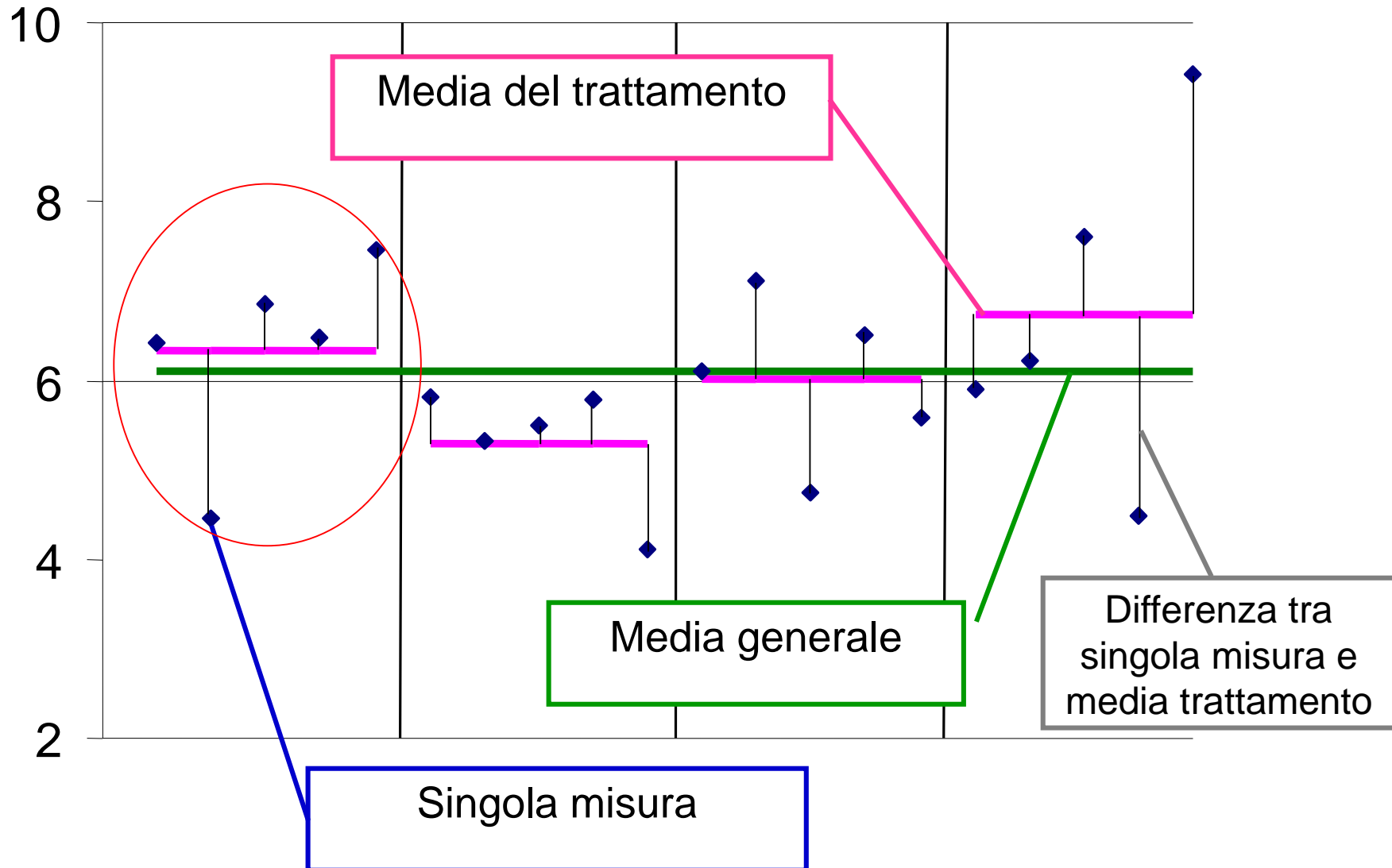
$$s^2 = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2}{\sum_{i=1}^a (n_i - 1)}$$

$a$  = numero di trattamenti = 1, 2, ...,  $i$

$n$  = numero di individui per trattamento = 1, 2, ...,  $j$  o ripetizioni

Essa rappresenta la variabilità della popolazione derivante da effetti non dovuti ai trattamenti. **Variabilità IN**terna ai trattamenti (dovuta ad altro es. microtopografia).

# ANOVA: Varianza IN



# ANOVA: Varianza TRA

Una seconda varianza della popolazione (Varianza tra) può essere calcolata a partire dall'errore standard elevato al quadrato, cioè partendo dalla distribuzione delle medie campionarie. Se non c'è effetto dei trattamenti essa sarà una stima di un errore standard derivante dalla stessa variabilità della popolazione.

Nell'ipotesi che in tutti i trattamenti siano stati prelevati lo stesso numero di campioni (osservazioni), si può scrivere:

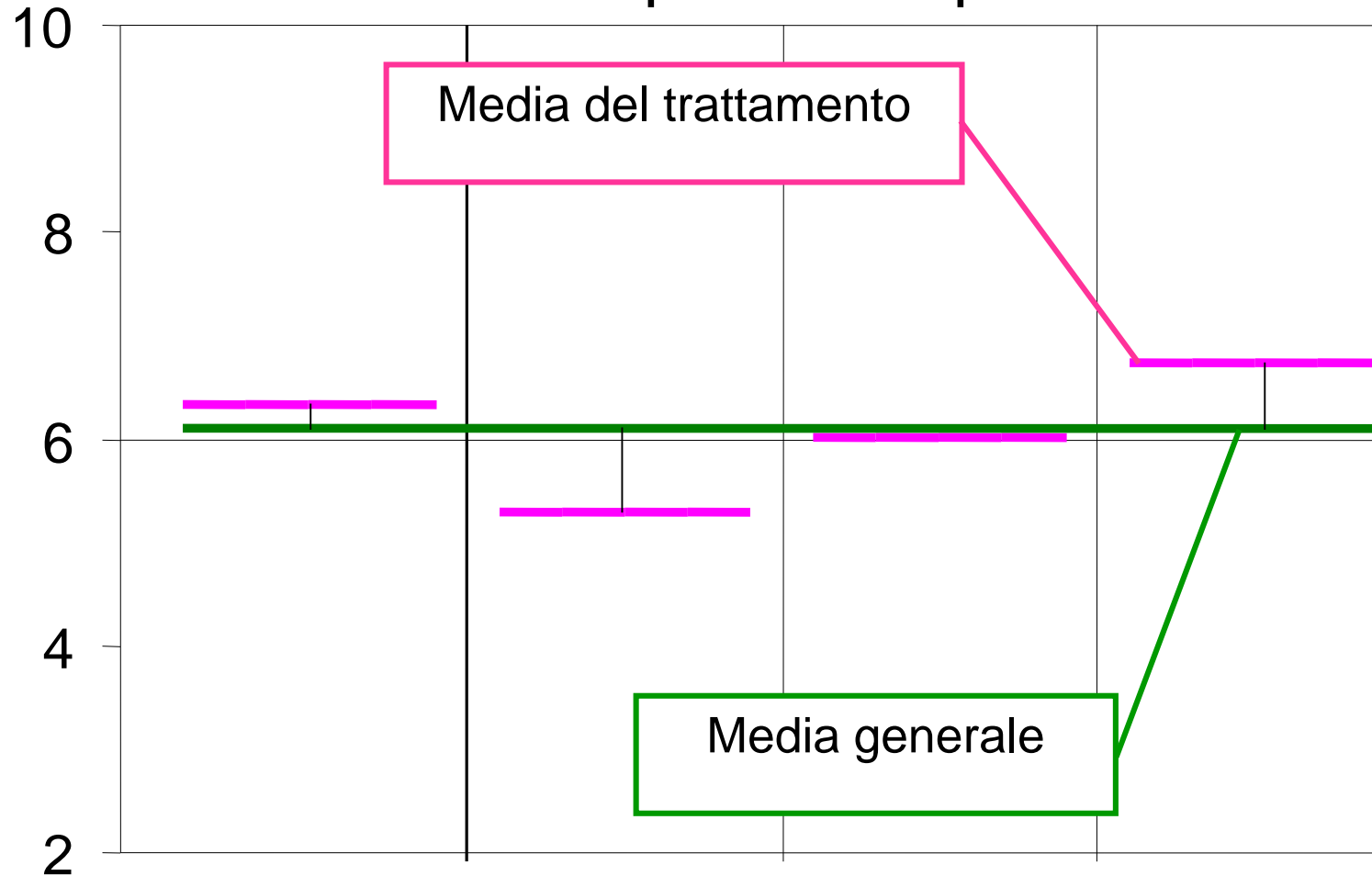
$$Err.st^2 = \frac{s^2}{n} = \frac{\sum_{i=1}^a (\bar{X}_i - \bar{\bar{X}})^2}{a-1} \qquad s^2 = \frac{\sum_{i=1}^a (\bar{X}_i - \bar{\bar{X}})^2}{a-1} \cdot n$$

Le due varianze **in** e **tra** saranno sicuramente differenti. Come verificare se le due varianze calcolate come stima della varianza della popolazione fanno parte della stessa popolazione o no?

## Il test F

# ANOVA: Varianza TRA

Errore standard TRA, che diventa varianza TRA se moltiplichiamo per n



# ANOVA: Il test F

$$F = \frac{s^2 tra}{s^2 in}$$

**Ipotesi nulla:  $H_0$ :**

$$s^2 tra = s^2 in$$

$$F = \frac{s^2 tra}{s^2 in} = 1$$

Se la stima della varianza tra trattamenti e quella all'interno dei trattamenti sono uguali, il rapporto risulta pari a 1.

Il rapporto difficilmente sarà = 1, ma se l'ipotesi nulla è vera, lo scostamento sarà solo dovuto al caso. È possibile calcolare per mezzo della distribuzione di  $F_{v(\text{num}), v(\text{den})}$  la probabilità associata al valore del rapporto ( $p$  di  $F$ ).

# ANOVA: Varianza totale

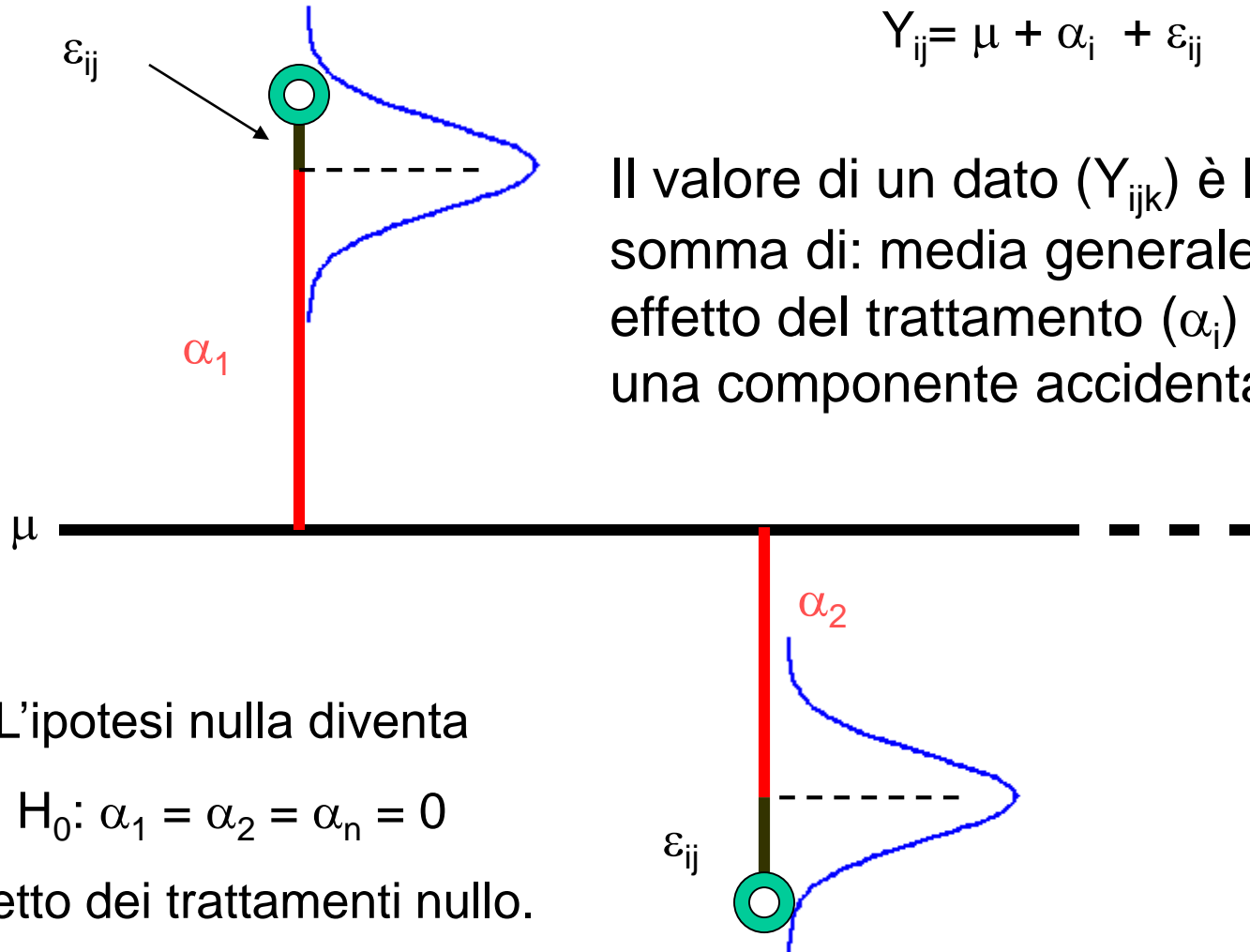
Varianza totale dalla media:

Si considerano tutti gli individui come appartenenti ad una unica popolazione

$$s^2 = \frac{\sum_{i=1}^a \sum_{j=1}^{n_i} (X_{ij} - \bar{\bar{X}})^2}{\left( \sum_{i=1}^a n_i \right) - 1}$$

Essa non serve ai fini del test F, ma può essere utile per il calcolo per differenza di componenti della varianza

## ANOVA: II MODELLO dell'ANOVA





# ANOVA: Assunzioni di validità

Le assunzioni di validità dipendono dal test F:

1. Gli errori  $\varepsilon$  devono essere tra loro indipendenti (randomizzazione del campionamento).
2. Devono essere distribuiti normalmente (deve essere verificata la normalità ed eventualmente i dati devono essere trasformati).
3. Le varianze fra i vari gruppi devono essere omogenee (occorre anche qui una verifica ed eventualmente una trasformazione).

# ANOVA:

## Test di normalità delle distribuzioni

Si raggruppano i dati in classi di ampiezza opportuna:

1. si calcolano le frequenze osservate in ogni classe
2. si calcolano le frequenze attese in ogni classe secondo una distribuzione normale
3. si testa per mezzo del  $\chi^2$  la bontà di adattamento delle frequenze attese alla distribuzione osservata.  $gdl = 2$ .

La normalità può essere testata visivamente utilizzando grafici QQ plot e matematicamente utilizzando le statistiche campionarie.

Praticamente vengono utilizzati i test:

- Shapiro-Wilk (molto potente per piccoli campioni)
- Kolmogorov-Smirnov (non parametrico, confronta le distribuzioni)

# ANOVA:

## Test per l'omogeneità delle varianze

### Test di Levene

Consiste nell'applicare l'analisi della varianza ad una via (one-way ANOVA) agli scarti in valore assoluto di ogni individuo dalla rispettiva media di trattamento.

Alcuni altri test invece di usare gli scarti in valore assoluto, usano gli scarti al quadrato ma sono molto meno utilizzati.

L'ipotesi nulla indica che le varianze dei vari gruppi sono tra loro omogenee.

$$H_0 = \sigma_a^2 = \sigma_b^2 = \sigma_c^2$$

Se viene rifiutata risulta che le varianze sono disomogenee e quindi i dati devono essere trasformati.

In questo caso sono contento se  $p > 0.05$

# ANOVA: La tabella Anova

	Devianza		Varianza		$p(F)$
<i>Origine della variazione</i>	<i>SQ</i>	<i>gdl</i>	<i>MQ</i>	<i>F</i>	<i>Valore di significatività</i>
Tra gruppi	5.364	3	1.788	1.180	0.349
In gruppi	24.248	16	1.516		
Totale	29.612	19			

**I gradi di libertà (gdl) sono:**

**Totale** =  $N - 1$  (numero dati rilevati - 1),

**Tra gruppi** =  $a - 1$  (numero trattamenti - 1),

**In gruppi** =  $a \cdot (n - 1)$  ((numero trattamenti \* (numerosità nel trattamento - 1)) nel caso in cui tutti i trattamenti abbiano lo stesso numero di individui.

# ANOVA: Confronti

Una volta verificato con il test ANOVA se esistono o no differenze significative fra le medie, occorre capire quali sono le medie o i gruppi di medie che effettivamente differiscono.

Per fare ciò abbiamo a disposizione due strumenti:

## **Confronti a priori o pianificati**

Sono delle ANOVA operate su sottogruppi di trattamenti. Sono più potenti di quelli non pianificati poiché operano su sottogruppi che non si sovrappongono mai, quindi non si traggono mai conclusioni combinate e non si devono mai calcolare probabilità combinate. La significatività è garantita. Permettono di verificare differenze fra medie come logicamente attese

## **Confronti a posteriori (detti Post-Hoc)**

Permettono di effettuare tutte le possibili combinazioni al fine di vedere quali medie hanno provocato la significatività.

Sono stati studiati differenti metodi per garantire la significatività prevista.

# ANOVA: Confronti a posteriori

Possono essere applicati solo dopo aver effettuato l'analisi della varianza.

Hanno le stesse condizioni di validità dell'ANOVA, ma l'allontanamento dalle assunzioni ha effetto differente sui vari test.

I test spesso sono del tipo confronto fra campioni indipendenti, ma gli autori hanno scelto soluzioni differenti per:

- garantire  $\alpha$  complessivo
- gestire il numero totale di confronti
- gestire l'inferenza simultanea

obiettivi:

- massima potenza sul singolo confronto
- conservazione  $\alpha$  complessivo

I test sono molteplici, ma possono essere ricondotti ad alcune categorie:

- test eseguiti sui confronti diretti 1 contro 1 (pair-wise)
- definizione di sottogruppi di trattamenti fra loro non differenti (experiment-wise)

# ANOVA: Test LSD *(pair-wise)*

Rappresenta l'intervallo fiduciale della differenza fra due medie calcolato con  $t$  di Student.

Soffre del problema della significatività complessiva che deriva dalla probabilità combinata (se i trattamenti sono pochi va bene, altrimenti no).

$$t_{n_1+n_2-2} \left( \frac{X_1 - X_2}{\sqrt{\frac{2}{n} \cdot s}} \right)$$

Dove:

$X_1$  e  $X_2$  sono le medie dei due trattamenti a confronto;

$s$  è la deviazione standard del residuo;

$n$  è il numero di casi per ogni trattamento che in un disegno bilanciato risulta essere uguale al numero di casi totali diviso il numero di trattamenti.

# ANOVA: Test Dunn Sidak *(pair-wise)*

Si tratta di un metodo uguale al test per due campioni indipendenti, ma con il valore di  $\alpha$  limite di ogni singolo confronto ridotto come segue

$$\alpha = 1 - (1 - \alpha_{\text{tot}})^{(1/\text{num. confronti})}$$

Su ogni singolo confronto infatti la protezione è  $(1 - \alpha_{\text{tot}})$  e quindi su  $m$  confronti  $(1 - \alpha_{\text{tot}})^m$ .

Ponendo che  $\alpha_{\text{tot}}$  deve essere uguale a 0.05, si deriva il valore di  $\alpha$  su ogni singolo confronto.



# ANOVA: Test Bonferroni *(pair-wise)*

Come principio risulta essere simile al test Dunn Sidak, ma Bonferroni fece una approssimazione che per un numero di confronti non elevato essa porta ad una valutazione di  $\alpha$  limite di ogni singolo confronto simile a quello calcolato con l'altro metodo:

$$\alpha = \alpha_{\text{tot}} / \text{numero confronti}$$

$$t_{n_1+n_2-2} \left( \frac{X_1 - X_2}{\sqrt{\frac{2}{n} \cdot s}} \right)$$

## ANOVA: Test Tukey *(pair-wise)*

La procedura è simile a quella basata su LSD, ma la distribuzione che viene considerata è quella di una variabile  $q$  definito intervallo di variazione «studentizzato» (studentized range).

$q$  dipende da  $\alpha$ , dal numero di trattamenti e dai gradi di libertà del residuo.

$$T_{\alpha} = q_{\alpha; p, nT-p} \cdot \frac{s}{\sqrt{n}}$$

## ANOVA: Test Dunnett

Basato sulla variabile  $t$  di Student testa tutti i trattamenti contro un controllo.

# ANOVA: Test SNK *(experiment-wise)*

Deriva dal test di Tukey: le medie vengono ordinate dalla minore alla maggiore. La differenza fra la media maggiore e la minore viene saggiata contro il valore limite dell'intervallo studentizzato secondo i seguenti parametri:

$\alpha = \alpha$  dell'ANOVA

$m = < p$  numero di medie

$$T_{\alpha} = q_{\alpha; m, nT-p} \cdot \frac{s}{\sqrt{n}}$$

Ad ogni confronto il valore di T limite cambia poiché cambia il valore m, ossia il numero di medie compreso nell'intervallo.

Se la differenza nell'intervallo è superiore al valore limite, si passa al secondo confronto in cui si esclude una media e si opera nuovamente il confronto. Al termine si avranno una serie di subset all'interno dei quali non esistono differenze significative.

# ANOVA una via: Esempio

Alimentazione dei pesci

Pesi degli animali al termine di 2 prove sperimentali



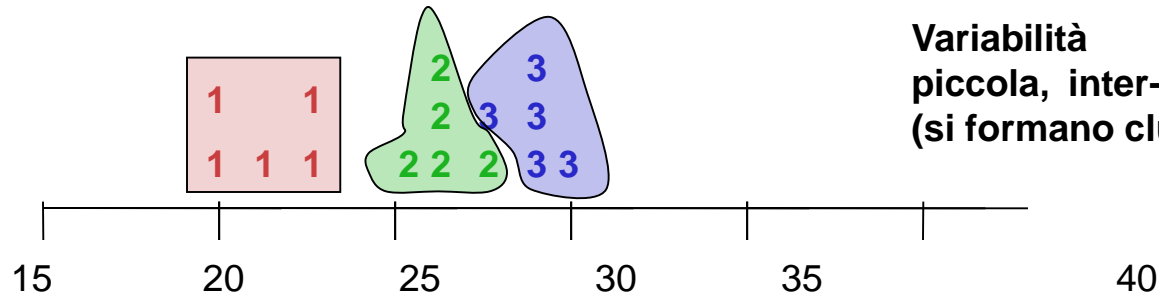
Prova 1			Prova 2		
Formula 1	Formula 2	Formula 3	Formula 1	Formula 2	Formula 3
20	25	28	18	27	17
22	27	28	24	20	37
21	26	27	17	29	29
22	26	29	22	31	21
20	26	28	24	23	36
<b>21</b>	<b>26</b>	<b>28</b>	<b>21</b>	<b>26</b>	<b>28</b>

Quale è la differenza di maggiore importanza nei dati?

# ANOVA una via: Esempio

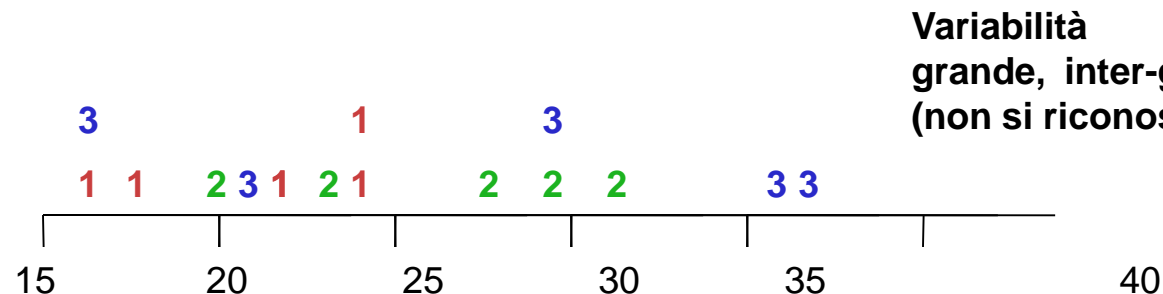
La differenza sta nel modo in cui sono distribuiti  
(I numeri colorati rappresentano le diverse formulazioni dell'alimento)

Prova 1



Variabilità intra-gruppo  
piccola, inter-gruppo grande  
(si formano clusters)

Prova 2



Variabilità intra-gruppo  
grande, inter-gruppo piccola  
(non si riconoscono clusters)

# ANOVA una via: Esempio

Per comparare le due prove  
(c'è differenza fra di esse?),  
formuliamo due ipotesi:

$H_0: \mu_1 = \mu_2 = \mu_3$       Non c'è differenza fra le formulazioni

$H_1: \mu_1 \neq \mu_2 \neq \mu_3$       C'è differenza fra le formulazioni

Effettuiamo 2 analisi della varianza  
ANOVA (prova1 e prova2)

# ANOVA una via: Esempio

## Piano sperimentale ed ANOVA

Esperimento completamente randomizzato:

- 1) 15 pesci
- 2) assegnazione casuale alla dieta

Si usa un'ANOVA a una via *one-way ANOVA*  
(o ad un fattore)

# ANOVA una via: Esempio

Variabilità complessiva ( $SS_T$ )



## Variabilità fra trattamenti

Deriva da:

1. Differenze fra soggetti
2. Errore sperimentale
3. Effetto dei trattamenti

## Variabilità nei trattamenti

Deriva da:

1. Differenze fra soggetti
2. Errore sperimentale

Si comparano con una statistica  $F$

$$F = \frac{\text{Variabilità fra trattamenti}}{\text{Variabilità nei trattamenti}} = \frac{SS_{\text{INTER}}}{SS_{\text{INTRA}}}$$

$$= \frac{\text{Effetto dei trattamenti} + \text{differenze fra soggetti} + \text{errore sperimentale}}{\text{differenze fra soggetti} + \text{errore sperimentale}}$$



# ANOVA una via: Esercizio 6.2

Considerando la prova 1...

Prova 1		
<b>Formula 1</b>	<b>Formula 2</b>	<b>Formula 3</b>
20	25	28
22	27	28
21	26	27
22	26	29
20	26	28
<b>21</b>	<b>26</b>	<b>28</b>

**Media generale = 25**

Come differiscono le diete?

Quale è la sorgente di variazione fra di esse?

# Esercizio 6.2

**Calcoli!**

1) Variabilità **complessiva**

= Somma dei quadrati totale =  $SS_T$

$$\begin{aligned}SS_T &= \sum X^2 - \frac{\sum(X)^2}{N} \\ &= 9513 - \frac{140,625}{15} \\ &= 138\end{aligned}$$

Prova 1		
Formula 1	Formula 2	Formula 3
20	25	28
22	27	28
21	26	27
22	26	29
20	26	28
21	26	28

**Media generale = 25**

Ma si può anche calcolare come:

$$\begin{aligned}SS_T &= (X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2 \\ &= (20 - 25)^2 + (22 - 25)^2 + (21 - 25)^2 + \dots + (28 - 25)^2 \\ &= 138\end{aligned}$$

Calcoli!

# Esercizio 6.2

2) Variabilità **intra**-gruppo

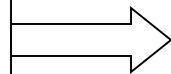
$$= \Sigma SS_{\text{intra}}$$

Prova 1		
Formula 1	Formula 2	Formula 3
20	25	28
22	27	28
21	26	27
22	26	29
20	26	28
21	26	28

$$\begin{aligned} SS_{\text{intra F1}} &= \Sigma X^2 - \frac{\Sigma(X)^2}{n} \\ &= 2209 - \frac{11025}{5} \\ &= 4 \end{aligned}$$

$$\begin{aligned} SS_{\text{intra F2}} &= 3382 - \frac{16900}{5} \\ &= 2 \end{aligned}$$

$$\begin{aligned} SS_{\text{intra F3}} &= 3922 - \frac{19600}{5} \\ &= 2 \end{aligned}$$



$$\begin{aligned} \Sigma SS &= SS_{\text{intra F1}} + SS_{\text{intra F2}} + SS_{\text{intra F3}} \\ &= 4 + 2 + 2 = 8 \end{aligned}$$

Media generale = 25

**Quindi  $SS_{\text{intra}} = 8$**

# Esercizio 6.2

## Calcoli!

3) Variabilità **inter**-gruppo,  
si calcola in due modi

$$SS_{\text{inter}} = \frac{\sum X^2}{n} - \frac{\sum (X)^2}{N}$$

3 (numero formule) 15 (numero pesci)

oppure

Si sottrae  $SS_{\text{intra}}$  da  $SS_{\text{T}}$

Poichè  $SS_{\text{T}} = SS_{\text{intra}} + SS_{\text{inter}}$

Quindi  $SS_{\text{inter}} = SS_{\text{T}} - SS_{\text{intra}} = 138 - 8 = 130$

Prova 1		
Formula 1	Formula 2	Formula 3
20	25	28
22	27	28
21	26	27
22	26	29
20	26	28
21	26	28

Media generale = 25

$$SS_{\text{inter}} = 130$$

Calcoli!

# Esercizio 6.2

4) Si calcola il Quadrato Medio

Ricordare la formula della varianza (con una piccola modifica)

$$s^2 = SS/gdl$$

Nell'ANOVA si sostituisce  $s^2$  col Quadrato Medio (MS):

$$MS_{inter} = \frac{SS_{inter}}{gdl_{inter}} = 130/2 = 65$$

$$MS_{intra} = \frac{SS_{intra}}{gdl_{intra}} = 8/12 = 0.66$$

$$\text{E quindi } F \text{ (la statistica dell'ANOVA)} = \frac{MS_{inter}}{MS_{intra}} = 65/0.66 = 97.59$$

# Esercizio 6.2

**Per una semplice ANOVA a una via**

I gradi di libertà sono :

$$\text{Gdl}_{\text{intra}} = N - K \rightarrow \text{Numero dati campione} - \text{Numero trattamenti} (15 - 3 = 12)$$

$$\text{Gdl}_{\text{inter}} = K - 1 \rightarrow \text{Numero trattamenti} - 1 (3 - 1 = 2)$$

$$\text{Gdl}_{\text{T}} = N - 1 \rightarrow \text{Numero dati campione} - 1 (15 - 1 = 14)$$

Dove N = numero dei dati totali (15 pesci)

K = numero dei trattamenti (3 formulazioni)

# Esercizio 6.2

I risultati dell'ANOVA si presentano in una tabella impostata così:

Sorgente di variazione	SS	gdl	MS (SS/gdl)	Stat F
Inter-gruppo (trattamento)	130	2 (K-1)	65	F = 97.59
Intra-gruppo (errore, residui)	8	12 (N-K)	0.65	
Totale	138	14 (N-1)		

# Esercizio 6.2

Il passo finale!

Si cerca il valore critico per  $p = 0.05$  e 2 e 12 gradi di libertà in una tavola di F

***Critical values of F for the 0.05 significance level:***

	1	2	3	4	5	6	7	8	9	10
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.39	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.97	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.10	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.28	3.06	2.91	2.80	2.71	2.64	2.59	2.54



# Esercizio 6.2

Il passo finale!

Si cerca il valore critico per  $p = .05$  e 2 e 12 gradi di libertà in una tavola di F

$$F(.05, 2, 12) = 3.89$$

Dato che il valore di F ottenuto (97.59) è molto maggiore di 3.89:

$$p \lll .05$$

Tornando alle ipotesi di partenza:

**$H_0: \mu_1 = \mu_2 = \mu_3$  Non c'è differenza fra le formulazioni**

**$H_1: \mu_1 \neq \mu_2 \neq \mu_3$  C'è differenza fra le formulazioni**

Si rigetta  $H_0$  e quindi si accetta  $H_1$ , cioè che esiste una differenza fra le formulazioni

# ANOVA

## Sommario dell'ANOVA

Variabilità totale

Fra trattamenti:

1. Differenze fra soggetti
2. Errore sperimentale
3. Effetto dei trattamenti

$$SS_T = \sum X^2 - \frac{\sum(X)^2}{N}$$

Nei trattamenti:

1. Differenze fra soggetti
2. Errore sperimentale

$$SS_{inter} = \frac{\sum X^2}{n} - \frac{\sum(X)^2}{N}$$

$$SS_{intra} = \sum SS_{INTRA Fx}$$

$$MS_{inter} = \frac{SS_{inter}}{gdl_{inter}}$$

$$MS_{intra} = \frac{SS_{intra}}{gdl_{intra}}$$

$$F = \frac{MS_{inter}}{MS_{intra}}$$