



Last updated 18 Aprile, 2023

Confronto tra due popolazioni

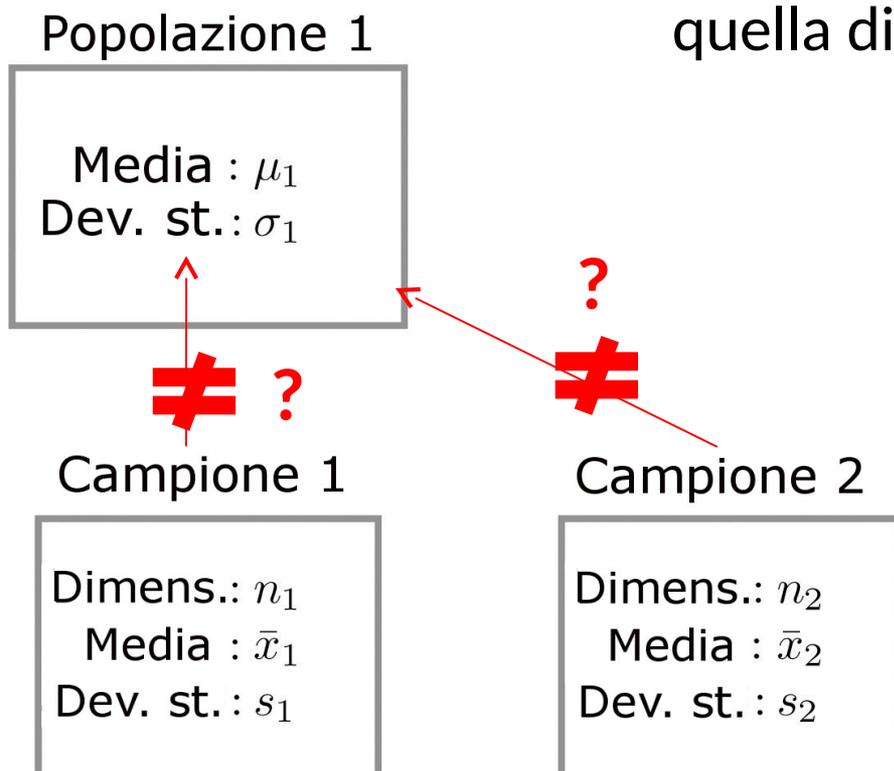
Lezione 6

G. Bacaro

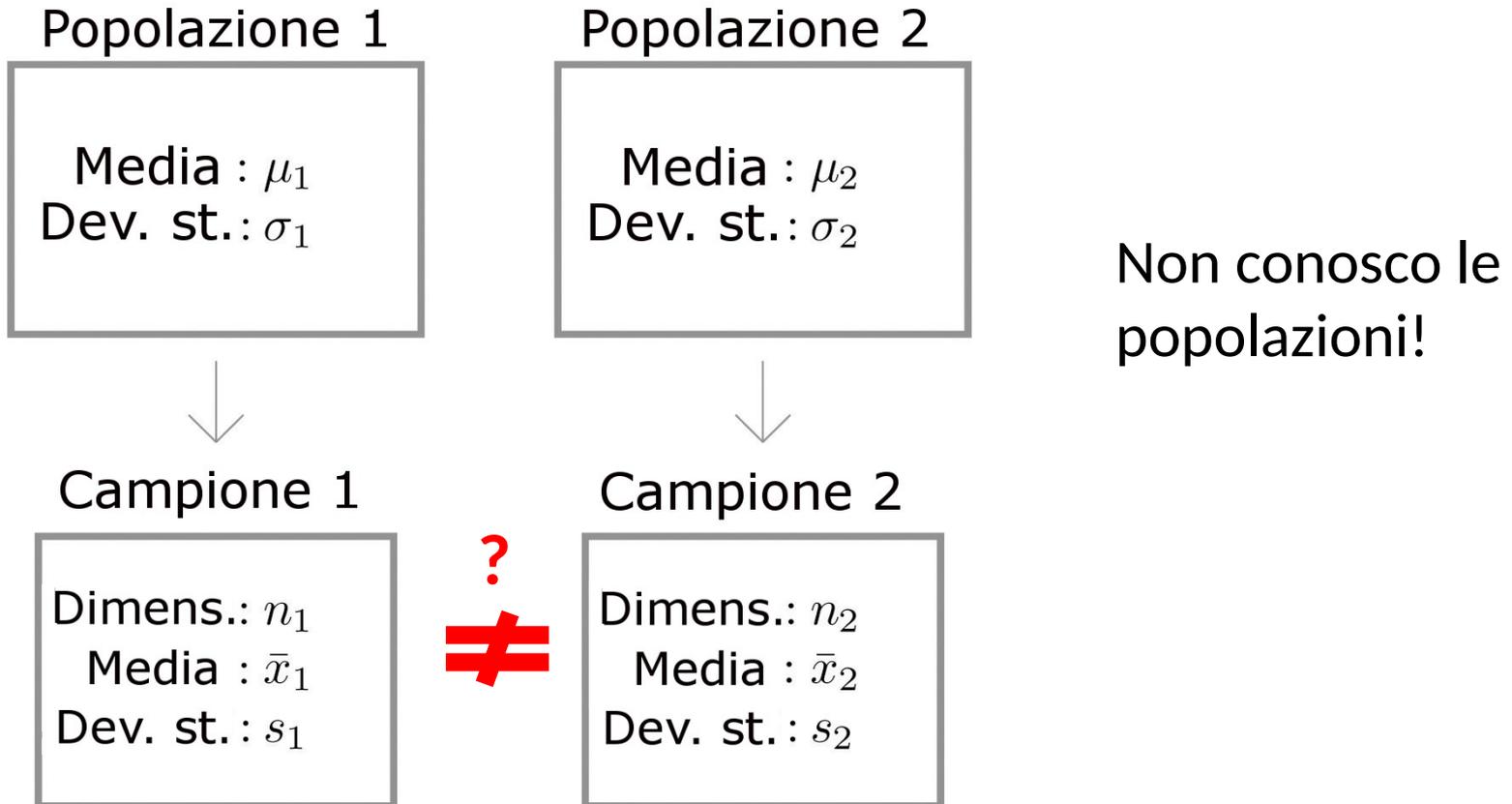
Statistica
CdL in Scienze e Tecnologie per l'Ambiente e la Natura
I anno, II semestre

Concetti visti nell'ultima lezione

Le media del campione è uguale e quella di una popolazione nota?



Confronto FRA due campioni

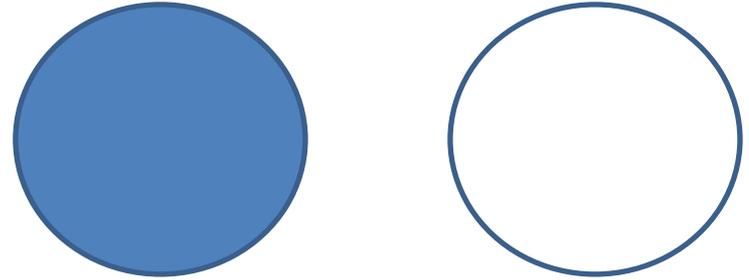


Il test t su due campioni

Ipotesi:

H_0 : le due medie sono uguali

H_a : le medie sono diverse ($\mu > \mu <$)

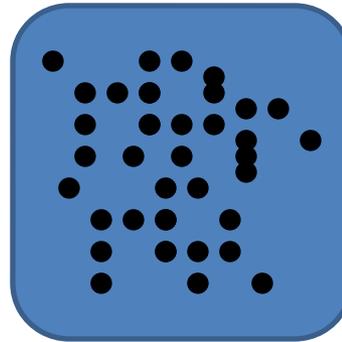
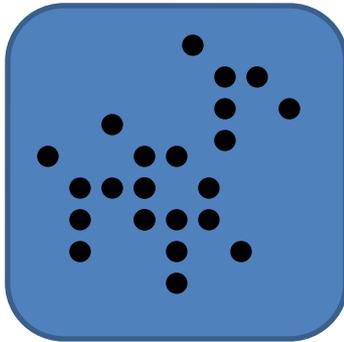


Assunzioni generali:

1. Indipendenza delle osservazioni (posso correggere per questo)
2. Normalità delle popolazioni a confronto
3. Omogeneità della varianza (posso correggere per questo)

1. Indipendenza delle osservazioni

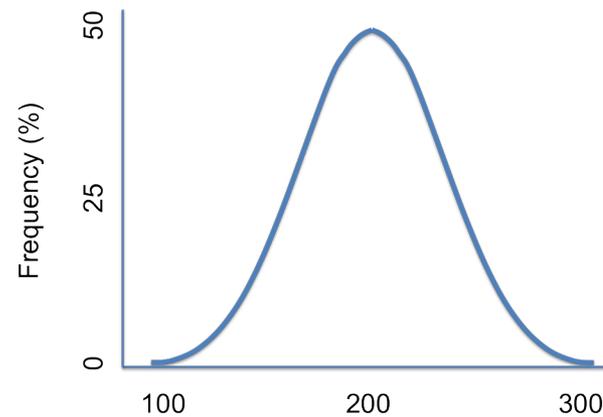
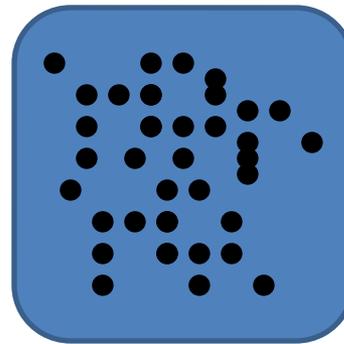
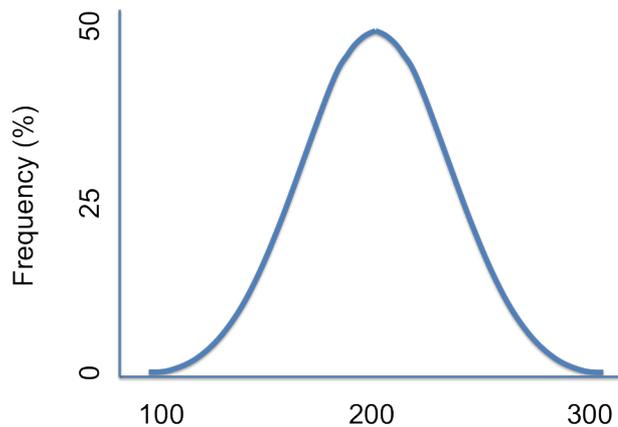
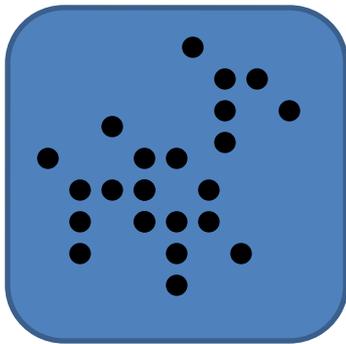
Ogni osservazione corrisponde ad una vera replica?



Attenzione al campionamento!!!

2. Normalità delle popolazioni a confronto

I due campioni devono provenire da popolazioni normali!



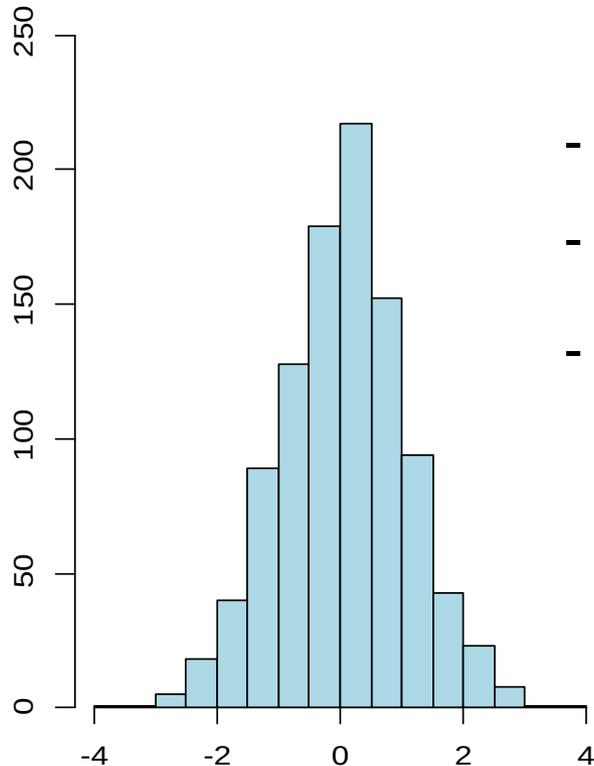
2. Normalità delle popolazioni a confronto

Operazioni per verificare la normalità

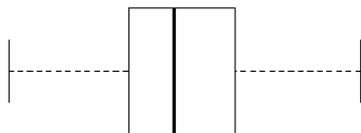
1. Confrontare le caratteristiche dei dati con quelle teoriche della distribuzione normale (es. mediana \approx media)
2. Analisi grafica (es. istogrammi)
3. Eseguire dei test (non considerati durante il corso)

2. Normalità delle popolazioni a confronto

Analisi dell'istogramma

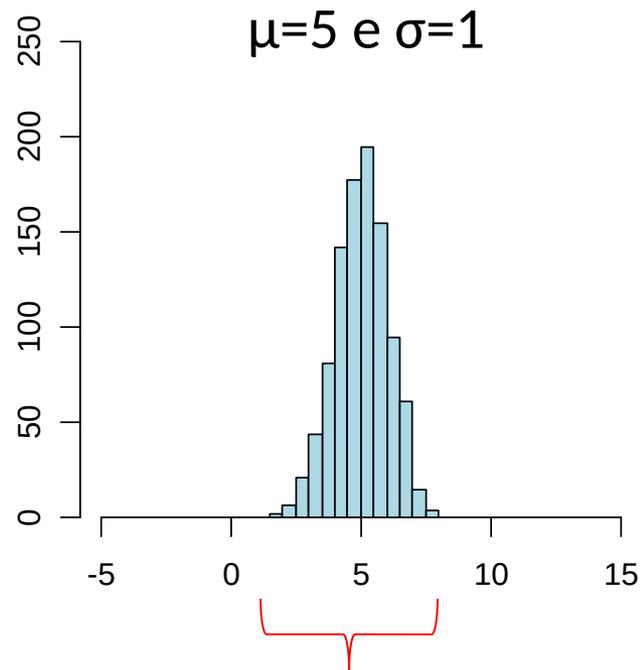
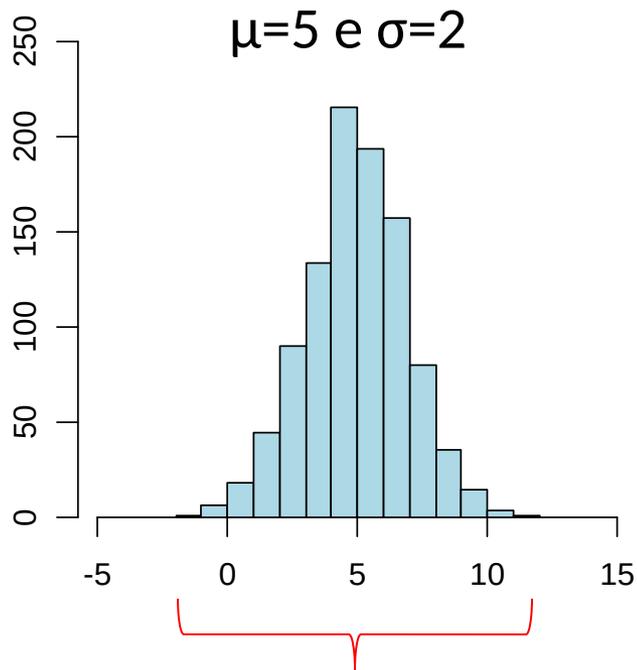


- Simmetria (media \approx mediana)
- c. 2/3 dei dati in un intervallo $\mu \pm \sigma$
- c. 95% dei dati in un intervallo $\mu \pm 2\sigma$



3. Omogeneità della varianza

Il livello di variabilità delle popolazioni a confronto deve essere simile!

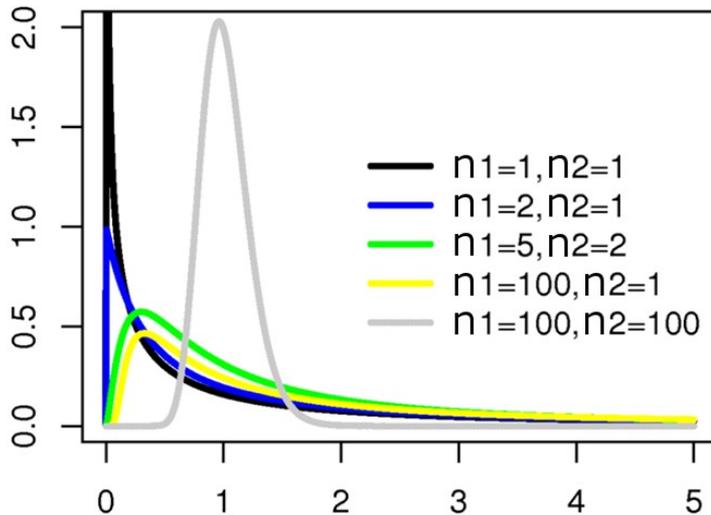


3. Omogeneità della varianza: Il test F

$$F = \frac{s_1^2}{s_2^2}$$

Varianza maggiore
Varianza minore

Distribuzione di probabilità che dipende dalla numerosità dei due campioni (n_1 e n_2)



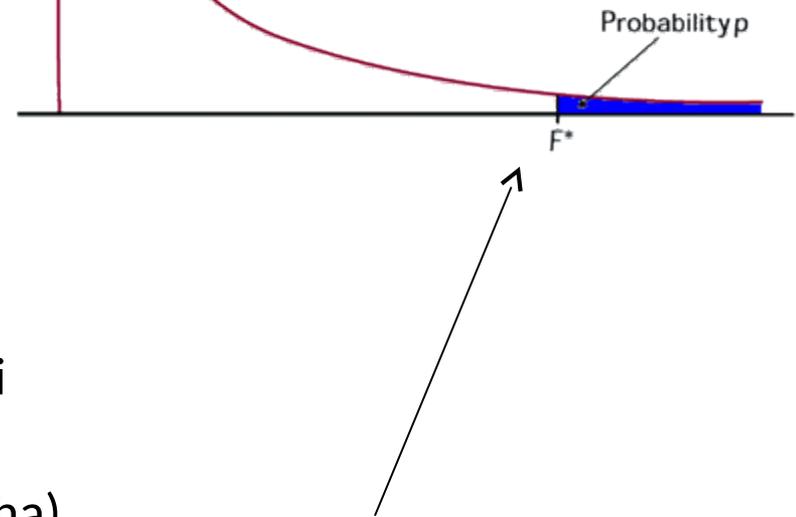
3. Omogeneità della varianza: Il test F

$$F_{\text{calcolato}} = \frac{s_1^2}{s_2^2} \quad \frac{\text{Varianza maggiore}}{\text{Varianza minore}}$$

H0: le due varianze sono uguali
Ha: le due varianze sono diverse

Test di ipotesi:

1. Calcolo la varianza dei due campioni
2. Determino il valore di $F_{\text{calcolato}}$
3. Decido il livello di significatività (alpha)
4. Determino il valore di F_{critico} (se la tavola dà P per **alpha/2**)
5. Se $F_{\text{calcolato}} > F_{\text{critico}}$ rifiuto H0
6. Conclusione: le varianze sono DIVERSE!



3. Omogeneità della varianza: Il test F

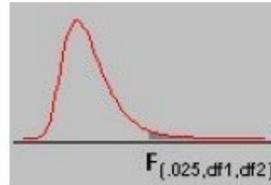
Numeratore: $n_1 - 1$

Denominatore: $n_2 - 1$



$F_{calcolato}$

La tavola dà un valore di F per una coda! Gli F qua sotto corrispondono a $\alpha=0.05$ a due code!

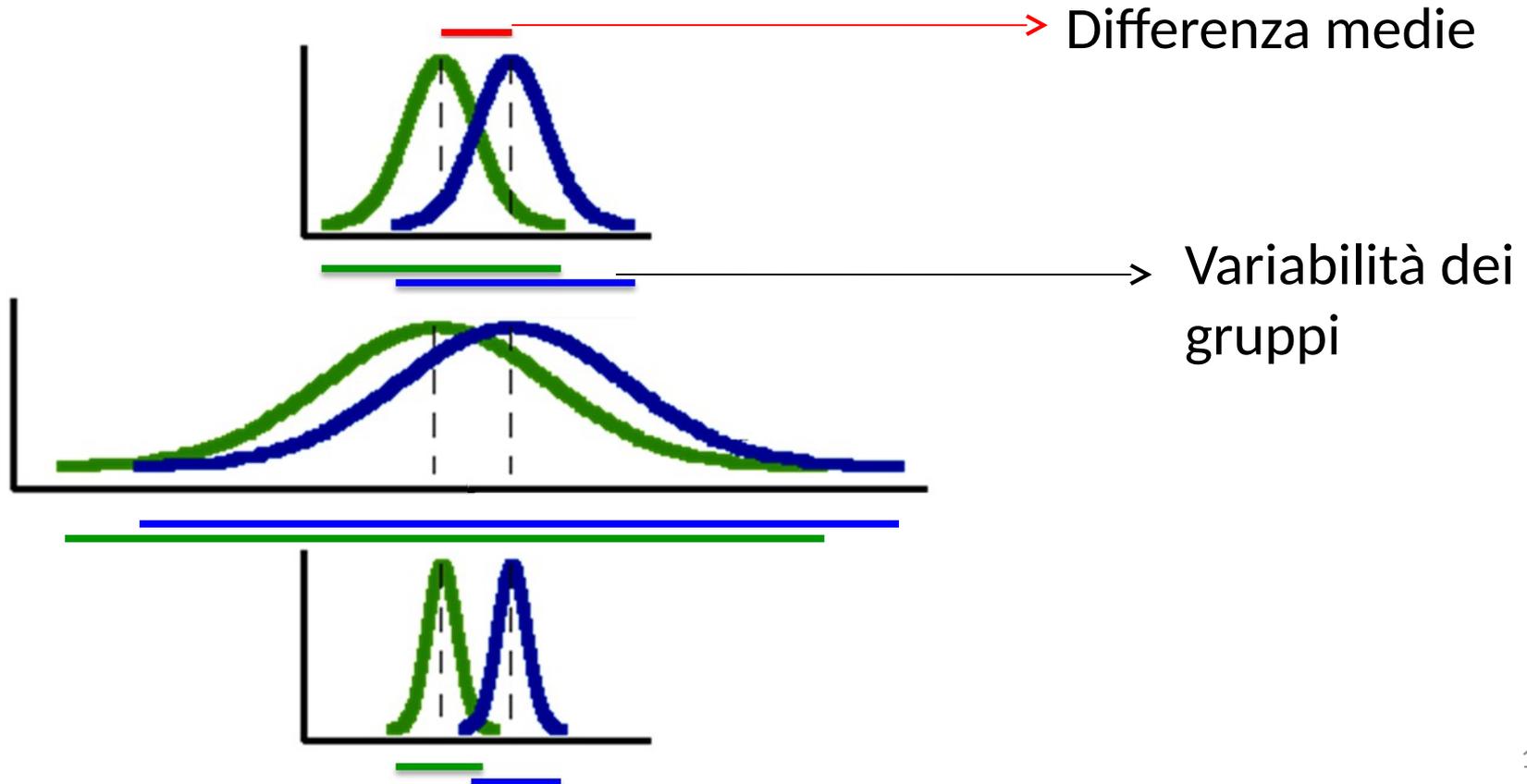


F Table for $\alpha= 0.025$ (1/3)

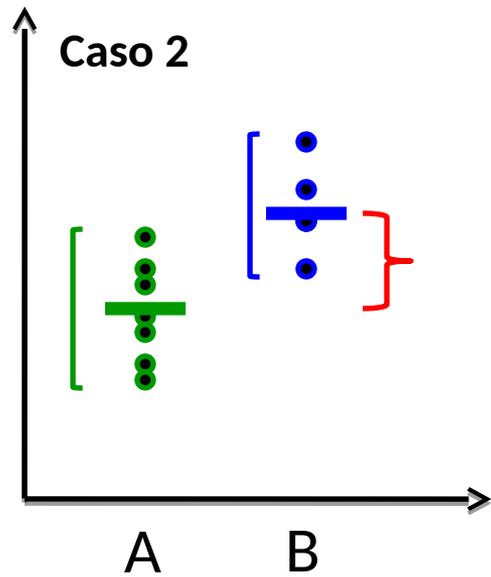
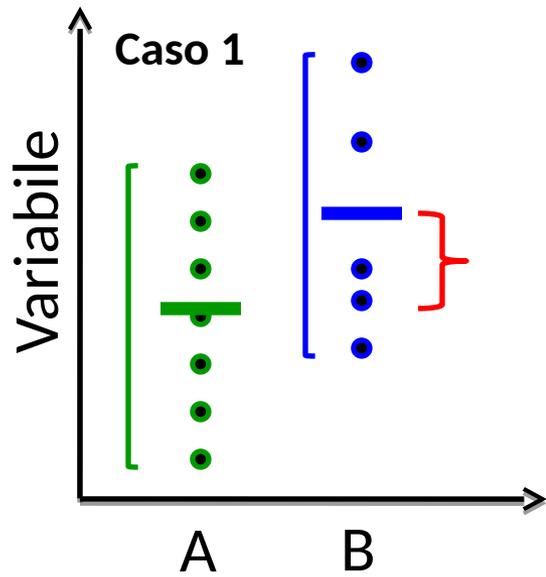
df2/df1	1	2	3	4	5	6	7	8	9
1	647.7890	799.5000	864.1630	899.5833	921.8479	937.1111	948.2169	956.6562	963.2846
2	38.5063	39.0000	39.1655	39.2484	39.2982	39.3315	39.3552	39.3730	39.3869
3	17.4434	16.0441	15.4392	15.1010	14.8848	14.7347	14.6244	14.5399	14.4731
4	12.2179	10.6491	9.9792	9.6045	9.3645	9.1973	9.0741	8.9796	8.9047
5	10.0070	8.4336	7.7636	7.3879	7.1464	6.9777	6.8531	6.7572	6.6811
6	8.8131	7.2599	6.5988	6.2272	5.9876	5.8198	5.6955	5.5996	5.5234
7	8.0727	6.5415	5.8898	5.5226	5.2852	5.1186	4.9949	4.8993	4.8232
8	7.5709	6.0595	5.4160	5.0526	4.8173	4.6517	4.5286	4.4333	4.3572
9	7.2093	5.7147	5.0781	4.7181	4.4844	4.3197	4.1970	4.1020	4.0260
10	6.9367	5.4564	4.8256	4.4683	4.2361	4.0721	3.9498	3.8549	3.7790
11	6.7241	5.2559	4.6300	4.2751	4.0440	3.8807	3.7586	3.6638	3.5879
12	6.5538	5.0959	4.4742	4.1212	3.8911	3.7283	3.6065	3.5118	3.4358

Il test t

$$t_{\text{calcolato}} = \frac{\text{Misura legata alla differenza fra le medie}}{\text{Misura di variabilità dentro i gruppi}}$$



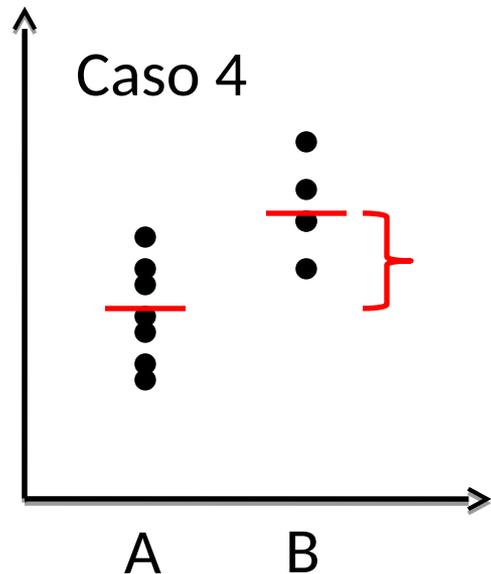
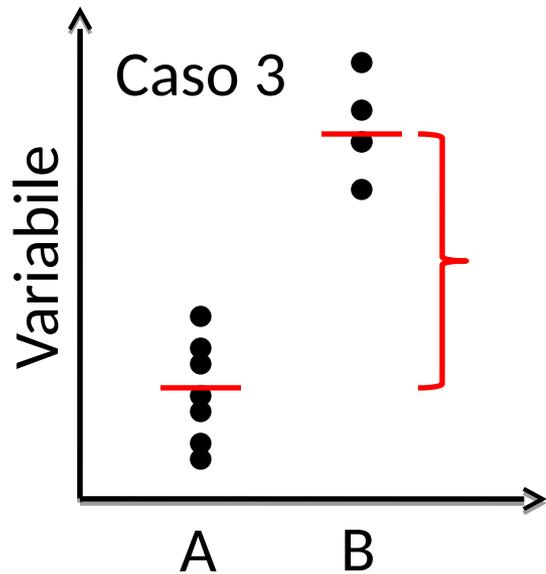
Il test t



} Differenza fra le medie

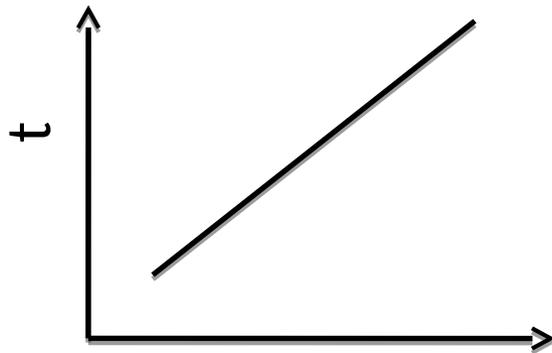
[Variabilità A

[Variabilità B

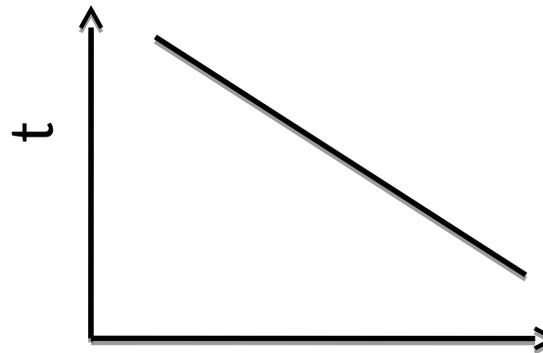


Il test t

$$t_{\text{calcolato}} = \frac{\text{Differenza fra le medie}}{\text{Errore standard della differenza}}$$



Differenza fra medie

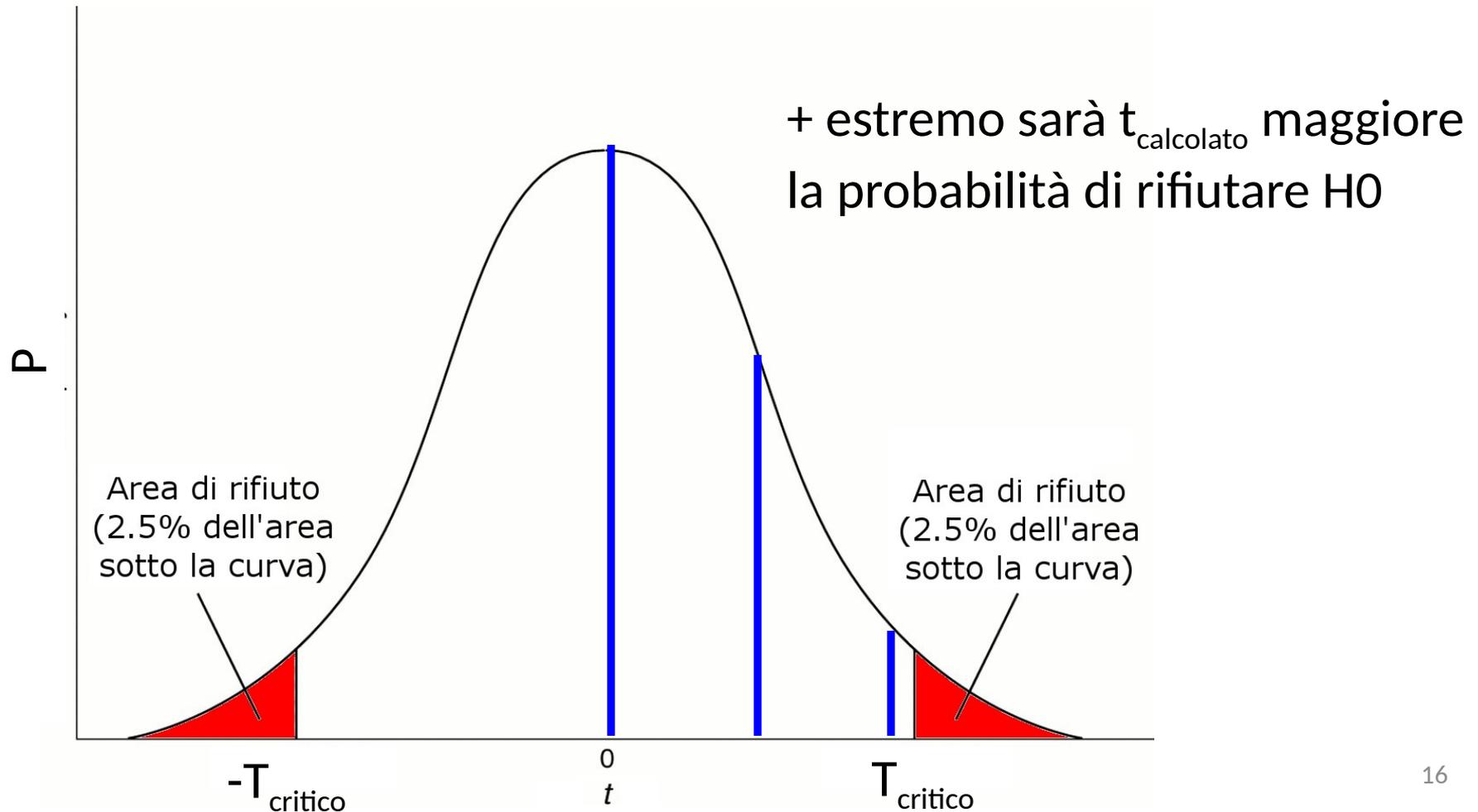


Variabilità dentro i gruppi

Più estremo sarà t calcolato maggiore sarà la probabilità di rifiutare H0

Il test t

$$t_{\text{calcolato}} = \frac{\text{Differenza fra le medie}}{\text{Errore standard della differenza}}$$



Come scegliere il test t giusto a partire dalle assunzioni

Indipendenza

NO

Test t appaiato

$$t = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}}$$

SÌ

Test t non appaiati

$$s_2^2 = s_1^2$$

Test t per
pop. omoschedastiche

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

$$s_2^2 \neq s_1^2$$

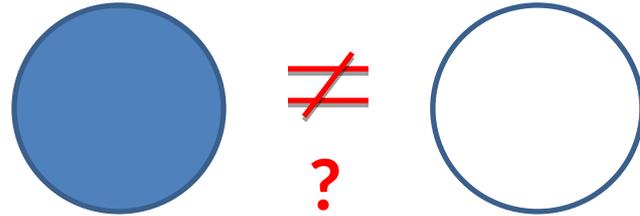
Test t per
pop. eteroschedastiche
Welch t-test
(formula complessa
richiesto un PC)

Campioni indipendenti omoschedastici: Test t!

$$t_{\text{calcolato}} = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$



$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$



Varianza combinata ("pooled")

I gradi di libertà sono $n_1 + n_2 - 2$ per T_{critico}

Campioni indipendenti omoschedastici: Test t!

H₀: le due medie sono uguali

H_a: le due medie sono diverse

Test di ipotesi:

1. Calcolo la varianza combinata dei due campioni
2. Determino il valore di $t_{\text{calcolato}}$
3. Decido il livello di significatività (alpha, 1 o 2 code?)
4. Determino il valore di t_{critico}
5. Se $|t_{\text{calcolato}}| > |t_{\text{critico}}|$ rifiuto H₀
6. Conclusione: le medie sono DIVERSE!

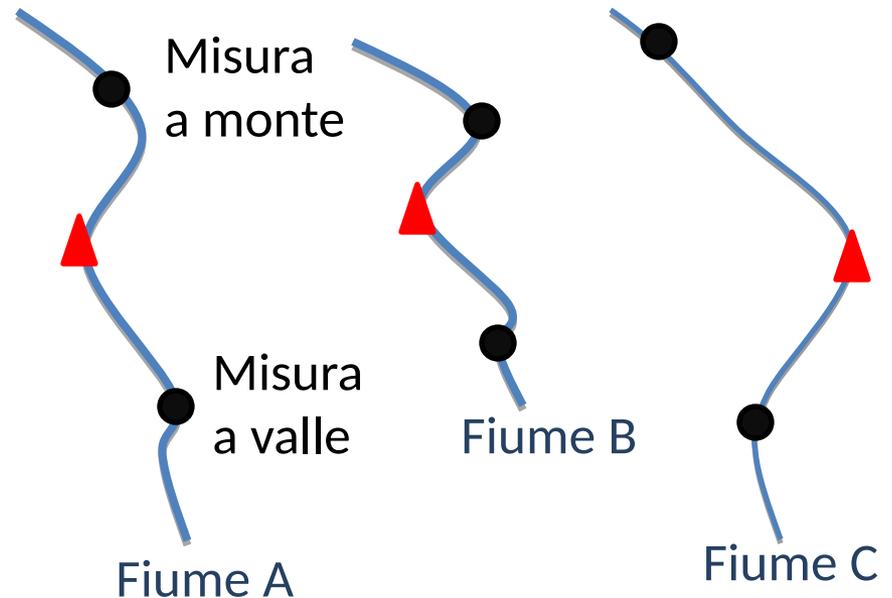
I gradi di libertà sono n_1+n_2-2 per T_{critico}

Campioni appaiati: 2 casi

1. Misure ripetute

Studente	Prima	Dopo
A	22	23
B	23	24
C	24	24
D	25	25
E	20	21
F	18	18
G	18	18
H	19	20

2. Correlazione nello spazio



▲ Industria tessile

● [Ammoniaca] in acqua

Campioni appaiati: Test t

$$t = \frac{\bar{D}}{\frac{S_D}{\sqrt{n}}}$$

$$\bar{D} = \frac{\sum D_i}{n}$$

Media delle differenze

$$S_D = \sqrt{\frac{\sum (D_i - \bar{D})^2}{n - 1}}$$

Deviazione standard delle differenze

n

Numero di coppie

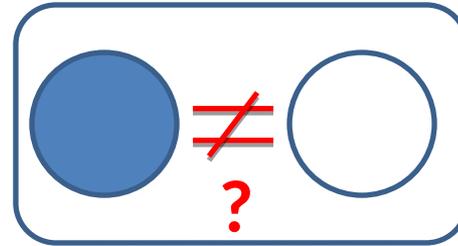
Studente	Prima	Dopo	D_i
A	22	23	1
B	23	24	1
C	24	24	0
D	25	25	0
E	20	21	1
F	18	18	0
G	18	18	0
H	19	20	1

I gradi di libertà sono $n-1$ per t_{critico}

Campioni appaiati: Test t

H₀: le due medie sono uguali

H_a: le due medie sono diverse



Test di ipotesi:

1. Determino il valore di $t_{\text{calcolato}}$
2. Decido il livello di significatività (alpha, 1 o 2 code?)
3. Determino il valore di t_{critico}
4. Se $|t_{\text{calcolato}}| > |t_{\text{critico}}|$ rifiuto H₀
5. Conclusione: le medie sono DIVERSE!

I gradi di libertà sono n-1 per t_{critico}