

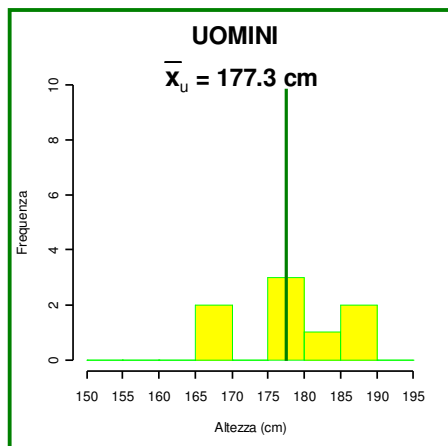
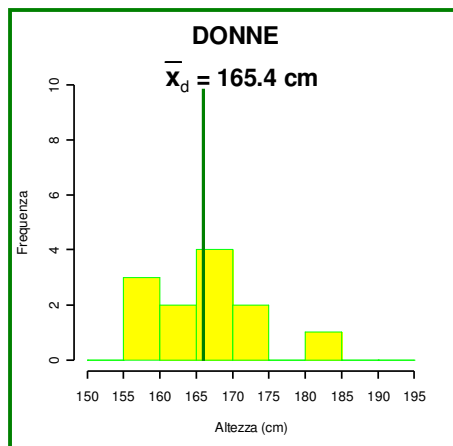
Metodi statistici per lo studio dei fenomeni biologici

Alla fine di questa lezione dovrete essere in grado di:

- ❑ descrivere la distribuzione di campionamento della differenza di due medie
- ❑ costruire gli intervalli di confidenza delle differenza di due medie con campioni indipendenti
- ❑ spiegare la logica del test di significatività statistica
- ❑ spiegare il concetto di ipotesi nulla
- ❑ spiegare il concetto di errore di I e II tipo
- ❑ effettuare un test di significatività per confrontare due medie di campioni indipendenti

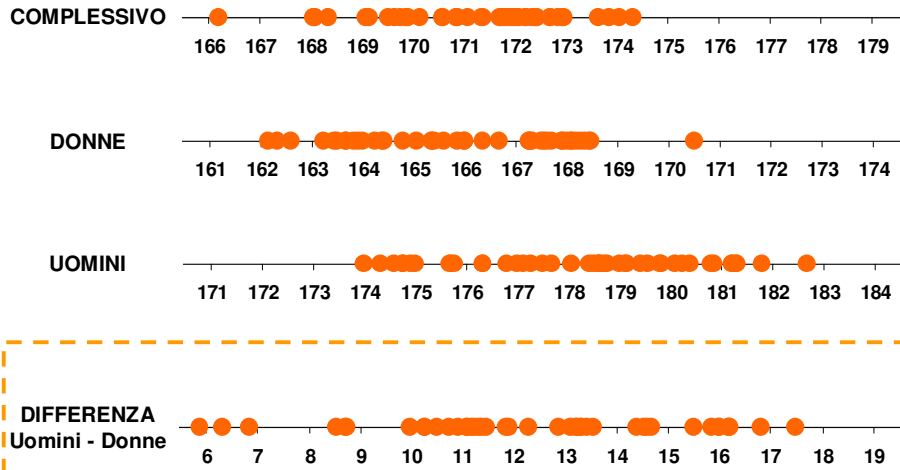


Quant'è la differenza di altezza fra uomini e donne?

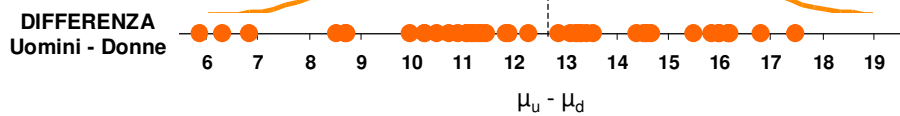


La differenza dell'altezza tra Uomini e Donne nel campione: $\bar{x}_u - \bar{x}_d = 11.8$ cm

Distribuzione delle Medie Campionarie dell'Altezza



Distribuzione delle Medie Campionarie dell'Altezza



Distribuzione delle Medie Campionarie

Caratteristiche della distribuzione delle medie campionarie

1. È approssimativamente Gaussiana
2. La media della distribuzione è μ
3. La deviazione standard della distribuzione è uguale a σ/\sqrt{n}

E la distribuzione della differenza delle medie campionarie?

Occorre distinguere in due casi differenti:

- ✓ Campioni Indipendenti (es. Uomini-Donne)
- ✓ Campioni Appaiati (es. 2 misure ripetute)

Campioni Indipendenti

Distribuzione della differenza delle medie campionarie

1. È approssimativamente Gaussiana
2. La media della distribuzione è $\mu_1 - \mu_2$
3. L'errore standard della distribuzione è uguale a:

$$\sigma \sqrt{\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)}$$

La varianza delle due popolazioni è uguale

Di questa situazione non ci occupiamo

$$\sqrt{\left(\frac{\sigma_1^2}{n_1}\right) + \left(\frac{\sigma_2^2}{n_2}\right)}$$

La varianza delle due popolazioni non è uguale

Calcolo dell'Intervallo di Confidenza al 95%

$$\left((\bar{X}_u - \bar{X}_d) - z \cdot \sigma \cdot \sqrt{\left(\frac{1}{n_u}\right) + \left(\frac{1}{n_d}\right)}, (\bar{X}_u - \bar{X}_d) + z \cdot \sigma \cdot \sqrt{\left(\frac{1}{n_u}\right) + \left(\frac{1}{n_d}\right)} \right)$$

Informazioni

$$n_u = 8$$

$$n_d = 12$$

$$\bar{x}_u = 177.3$$

$$\bar{x}_d = 165.4$$

$$\sigma = 8.5$$

$$z = 1.96$$

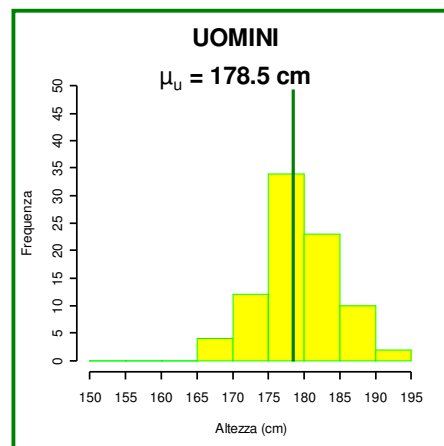
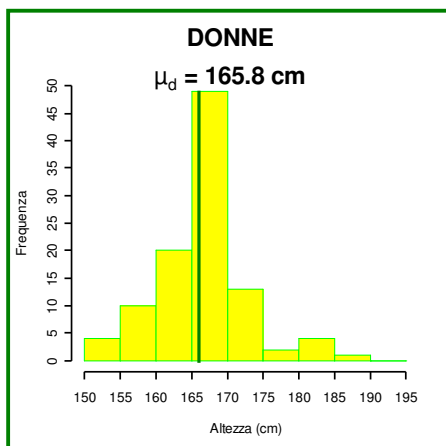
Limite Inferiore

$$(177.3 - 165.4) - 1.96 \cdot 8.5 \cdot \sqrt{\left(\frac{1}{8}\right) + \left(\frac{1}{12}\right)} = 11.9 - 7.6 = 4.3$$

Limite Superiore

$$(177.3 - 165.4) + 1.96 \cdot 8.5 \cdot \sqrt{\left(\frac{1}{8}\right) + \left(\frac{1}{12}\right)} = 11.9 + 7.6 = 19.5$$

Altezza della Popolazione di Studenti per Genere



La differenza dell'altezza tra Uomini e Donne nella popolazione: $\mu_u - \mu_d = 12.7 \text{ cm}$

Se σ è sconosciuta ?

Problema

Se la varianza della popolazione σ^2 non è nota ?

Soluzione

Utilizzo le due varianze campionarie s_1^2 e s_2^2 per stimare la varianza della popolazione:

$$S^2_{pooled} = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}$$

In questo caso la distribuzione delle differenza delle medie approssima alla distribuzione t con $n_1 + n_2 - 2$ gradi di libertà

Calcolo dell'Intervallo di Confidenza al 95%

$$\left((\bar{X}_u - \bar{X}_d) - t_{gl} \cdot S_{pooled} \cdot \sqrt{\left(\frac{1}{n_u}\right) + \left(\frac{1}{n_d}\right)}, (\bar{X}_u - \bar{X}_d) + t_{gl} \cdot S_{pooled} \cdot \sqrt{\left(\frac{1}{n_u}\right) + \left(\frac{1}{n_d}\right)} \right)$$

Informazioni

$$n_u = 8$$

$$n_d = 12$$

$$\bar{x}_u = 177.3$$

$$\bar{x}_d = 165.4$$

$$s_u^2 = 58.8$$

$$s_d^2 = 51.5$$

$$gl = 18$$

$$t_{18} = 2.10$$

Limite Inferiore

$$(177.3 - 165.4) - 2.10 \cdot 7.4 \cdot \sqrt{\left(\frac{1}{8}\right) + \left(\frac{1}{12}\right)} = 11.9 - 7.1 = 4.8$$

Limite Superiore

$$(177.3 - 165.4) + 2.10 \cdot 7.4 \cdot \sqrt{\left(\frac{1}{8}\right) + \left(\frac{1}{12}\right)} = 11.9 + 7.1 = 19.0$$

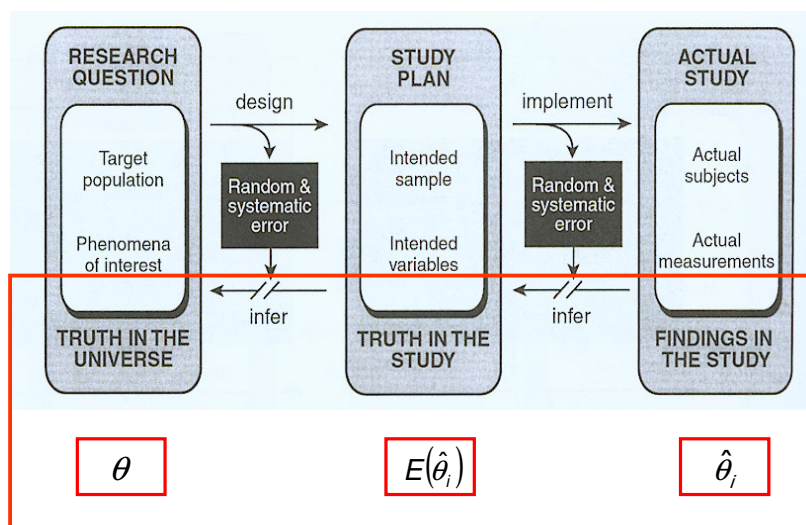
Il Test Statistico d'Ipotesi

I soggetti con Rh+ hanno una altezza differente rispetto ai soggetti con Rh-?

Lo studio verrà svolto su un campione di 20 studenti del primo anno di medicina

Approccio ipotetico-deduttivo

La struttura della ricerca clinica



Il Test Statistico d'Ipotesi

Fasi logiche

Ipotesi iniziale

Risultati attesi
sotto H_0

Confronto
osservati-attesi

Conclusioni

Fasi operative

Ipotesi nulla (H_0)

Distribuzione di
campionamento

Test statistico

Rifiuto/non
rifiuto di H_0

Il Test Statistico d'Ipotesi

In tribunale ...

L'imputato è presunto innocente ...

La colpevolezza va dimostrata oltre ogni ...

Ragionevole dubbio in un ...

Dibattimento, alla fine del quale si può ...

Condannare ...

... un colpevole



Giustizia è fatta!

... un innocente



Errore giudiziario!

Assolvere

... un innocente



Giustizia è fatta!

... un colpevole



Errore giudiziario!

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Ipotesi iniziale

Ipotesi nulla (H_0)

L'ipotesi nulla H_0 è un'affermazione sull'effetto vero del trattamento che lo studio si propone di confutare. È considerata valida fino a prova contraria.

Se l'obiettivo è riconoscere un'eventuale differenza tra i gruppi, l'ipotesi nulla è che le medie dei due gruppi siano uguali.

$H_0: \mu_{Rh+} = \mu_{Rh-}$ oppure $\mu_{Rh+} - \mu_{Rh-} = 0$ *ipotesi nulla*

$H_1: \mu_{Rh+} \neq \mu_{Rh-}$ oppure $\mu_{Rh+} - \mu_{Rh-} \neq 0$ *ipotesi alternativa*

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Risultati attesi sotto H_0

Distribuzione di campionamento

Distribuzione delle differenze campionarie teoricamente possibili se i gruppi fossero uguali.

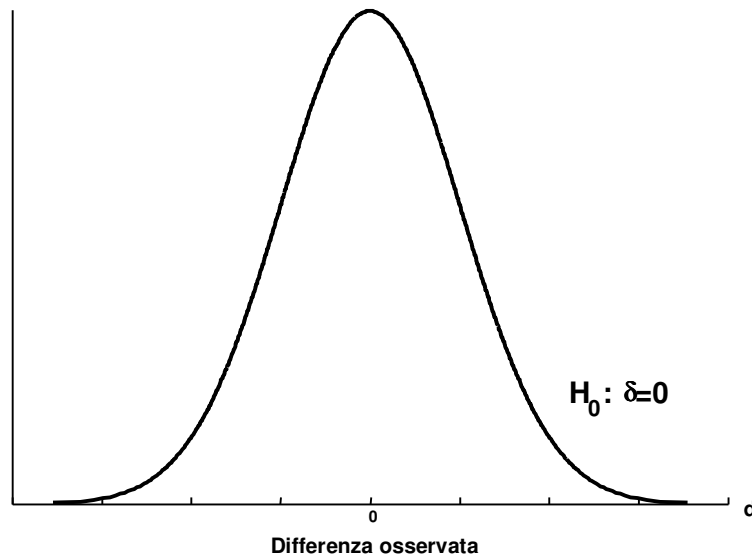
1. È approssimativamente Gaussiana
2. La media della distribuzione è 0
3. L'errore standard della distribuzione è uguale a:

$$\sigma \sqrt{\left(\frac{1}{n_1}\right) + \left(\frac{1}{n_2}\right)}$$

Definita da H_0

La varianza delle due popolazioni è uguale

Distribuzione di campionamento se H_0 fosse vera



Il Test Statistico d'Ipotesi

Fasi logiche

**Confronto
osservati-attesi**

Fasi operative

Test statistico

- Si valuta la distanza" tra risultato campionario e teorico atteso
- Si calcola la plausibilità di H_0 visti i dati
 - Quanto è probabile che la differenza effettivamente osservata sia imputabile al caso (se non vi sono differenze fra i gruppi)?
 - maggiore è la distanza del risultato osservato dall'ipotesi nulla, minore è la probabilità che il risultato osservato possa essere casuale

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Confronto
osservati-attesi

Test statistico

Il campione di studenti ha dato i seguenti risultati:

$$n_+ = 15$$

$$n_- = 5$$

$$\bar{x}_+ = 171.3$$

$$\bar{x}_- = 166.6$$

La differenza tra le due medie è pari a 4.7 cm

Quanto è probabile che questa differenza sia imputabile al caso (se in realtà l'altezza media è uguale nei due gruppi)?

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Confronto
osservati-attesi

Test statistico

Questo **Test** in realtà già lo conosciamo:

$$Z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Sotto l'ipotesi H_0 la differenza è nulla

La deviazione standard
potremmo non conoscerla, ma
sappiamo come stimarla

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Confronto
osservati-attesi

Test statistico

Modifichiamo il **Test** nel caso σ non sia nota:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \longrightarrow \quad t_{gl} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{s_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Il Test Statistico d'Ipotesi

Fasi logiche

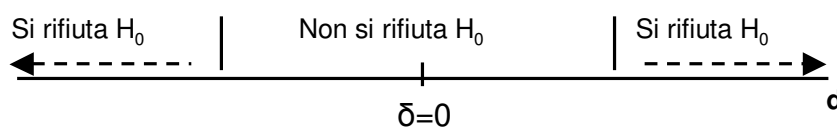
Fasi operative

Conclusioni

Rifiuto/non
rifiuto di H_0

- Serve una regola che consenta di rifiutare H_0 se i dati campionari non sono "consistenti" con H_0
- Si rifiuta H_0 se d è molto più piccola o molto più grande di zero: ma quanto più grande o più piccolo?

SI DEVE SCEGLIERE UNA REGIONE CRITICA

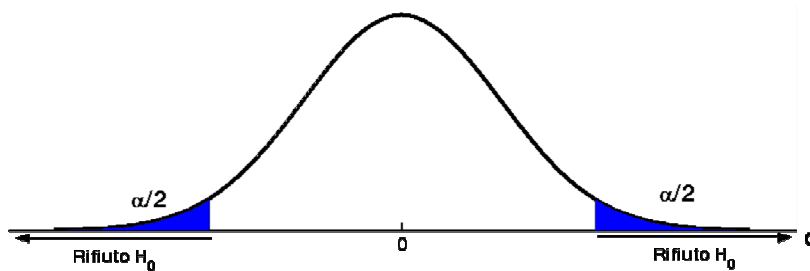


Distribuzione di campionamento sotto H_0

Regione critica di rifiuto (2 code)

$H_0: \delta = 0$

α = livello di
significatività



Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Conclusioni

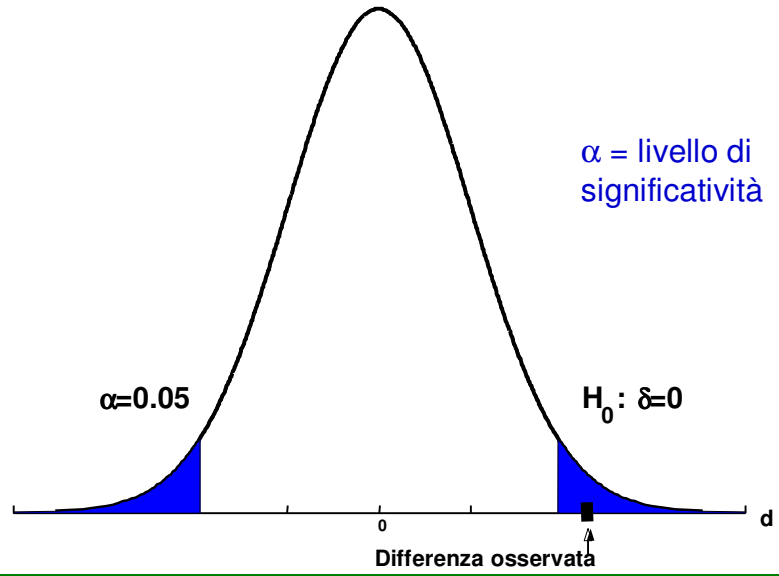
Rifiuto/non
rifiuto di H_0

Con il **Test Statistico** calcoliamo *la probabilità* che la differenza osservata sia imputabile al caso (se in realtà l'altezza media è uguale nei due gruppi)

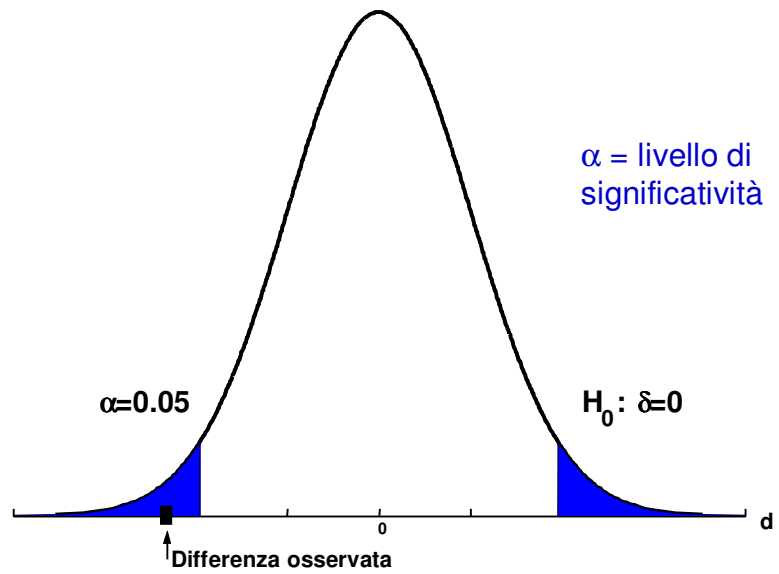
Se questa probabilità è 'piccola', ovvero se il risultato osservato è sufficientemente diverso da zero, il risultato si dice **statisticamente significativo**: abbiamo prove sufficienti per concludere che l'ipotesi nulla di assenza di efficacia sia falsa.

Errore associato: risultato falso positivo o di I tipo

Risultato statisticamente significativo



Risultato statisticamente significativo



Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Conclusioni

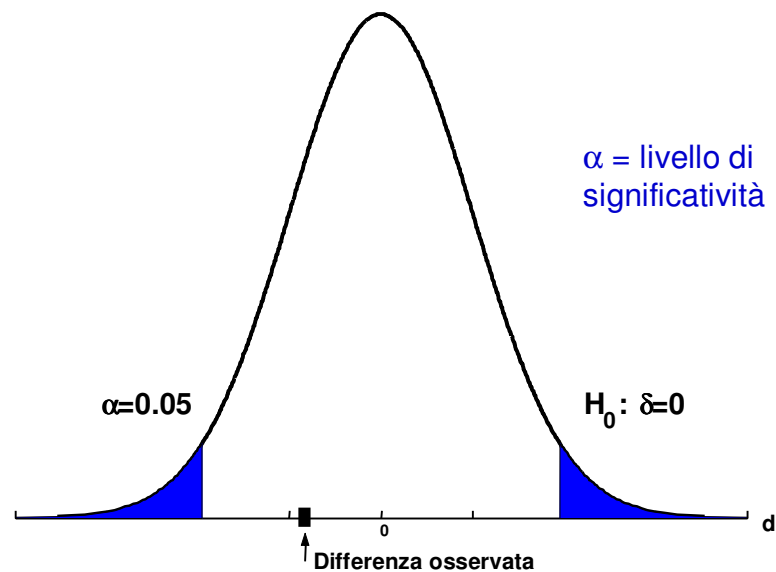
**Rifiuto/non
rifiuto di H_0**

Con il **Test Statistico** calcoliamo *la probabilità* che la differenza osservata sia imputabile al caso (se in realtà l'altezza media è uguale nei due gruppi)

Se questa probabilità non è piccola, ovvero se il risultato osservato non è sufficientemente diverso da zero, il risultato si dice **statisticamente non significativo**: non abbiamo, cioè, prove sufficienti per confutare l'ipotesi nulla di assenza di efficacia

Errore associato: risultato falso negativo o di II tipo

Risultato statisticamente non significativo



Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Confronto
osservati-attesi

Test statistico

Il campione di studenti ha dato i seguenti risultati:

$$n_+ = 15$$

$$n_- = 5$$

$$\bar{x}_+ = 171.3$$

$$\bar{x}_- = 166.6$$

La differenza tra le due medie è pari a 4.7 cm

Quanto è probabile che questa differenza sia imputabile al caso (se in realtà l'altezza media è uguale nei due gruppi)?

Il Test Statistico d'Ipotesi

Fasi logiche

Fasi operative

Confronto
osservati-attesi

Test statistico

Applichiamo il **Test** ai nostri dati:

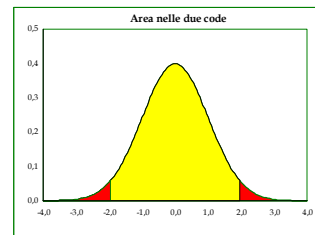
$$t_{gl} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$



$$t_{18} = \frac{(171.3 - 166.6) - (0)}{9.3 \cdot \sqrt{\frac{1}{15} + \frac{1}{5}}} = \frac{4.7}{4.8} = 0.98$$

Percentili della distribuzione t di Student

GL	PROBABILITA' (2 code)				PROBABILITA' (1 coda)			
	0,1	0,05	0,02	0,01	0,05	0,025	0,01	0,005
1	6,31	12,71	11,82	63,66	6,31	12,71	31,82	63,66
2	2,92	4,30	6,96	9,92	2,92	4,30	6,96	9,92
3	2,35	3,18	4,54	5,84	2,35	3,18	4,54	5,84
4	2,13	2,78	3,75	4,60	2,13	2,78	3,75	4,60
5	2,02	2,57	3,36	4,03	2,02	2,57	3,36	4,03
6	1,94	2,45	3,14	3,71	1,94	2,45	3,14	3,71
7	1,89	2,36	3,00	3,50	1,89	2,36	3,00	3,50
8	1,86	2,31	2,90	3,36	1,86	2,31	2,90	3,36
9	1,83	2,26	2,82	3,25	1,83	2,26	2,82	3,25
10	1,81	2,23	2,76	3,17	1,81	2,23	2,76	3,17
11	1,80	2,20	2,72	3,11	1,80	2,20	2,72	3,11
12	1,78	2,18	2,68	3,05	1,78	2,18	2,68	3,05
13	1,77	2,16	2,65	3,01	1,77	2,16	2,65	3,01
14	1,76	2,14	2,62	2,98	1,76	2,14	2,62	2,98
15	1,75	2,13	2,60	2,95	1,75	2,13	2,60	2,95
16	1,75	2,12	2,58	2,92	1,75	2,12	2,58	2,92
17	1,74	2,11	2,57	2,90	1,74	2,11	2,57	2,90
18	1,73	2,10	2,55	2,88	1,73	2,10	2,55	2,88
19	1,73	2,09	2,54	2,86	1,73	2,09	2,54	2,86
20	1,72	2,09	2,53	2,85	1,72	2,09	2,53	2,85
21	1,72	2,08	2,52	2,83	1,72	2,08	2,52	2,83
22	1,72	2,07	2,51	2,82	1,72	2,07	2,51	2,82
23	1,71	2,07	2,50	2,81	1,71	2,07	2,50	2,81
24	1,71	2,06	2,49	2,80	1,71	2,06	2,49	2,80
25	1,71	2,06	2,49	2,79	1,71	2,06	2,49	2,79
26	1,71	2,06	2,48	2,78	1,71	2,06	2,48	2,78
27	1,70	2,05	2,47	2,77	1,70	2,05	2,47	2,77
28	1,70	2,05	2,47	2,76	1,70	2,05	2,47	2,76
29	1,70	2,05	2,46	2,76	1,70	2,05	2,46	2,76
30	1,70	2,04	2,46	2,75	1,70	2,04	2,46	2,75
∞	1,64	1,96	2,05	2,33	1,64	1,96	2,05	2,33



Se $t < -2.10$ rifiuto H_0

Se $t > +2.10$ rifiuto H_0

Se $-2.10 < t < 2.10$ non rifiuto H_0

Il Test Statistico d'Ipotesi

Fasi logiche

Conclusioni

Fasi operative

Rifiuto/non rifiuto di H_0

