NOTE DALLE LEZIONI
DI
STATISTICA MEDICA
ED ESERCIZI

I METODI PER IL CONFRONTO DI MEDIE
(*Campioni non indipendenti)*

IL PROBLEMA

Si vuole verificare l'efficacia di una dieta sulla riduzione dei livelli di colesterolo in un gruppo di soggetti obesi.
Pertanto il livello di colesterolo è stato misurato prima e dopo la dieta

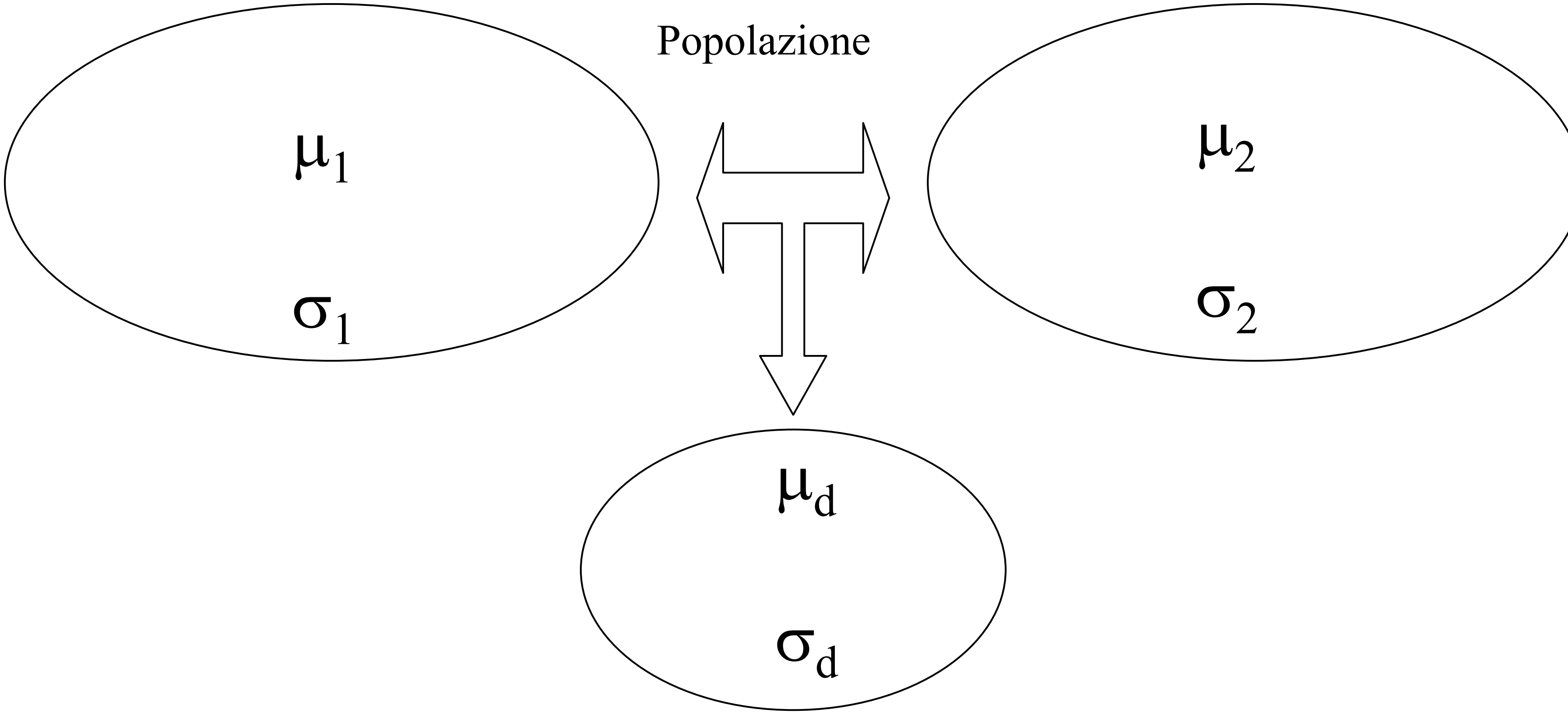
X_1 Prima	X_2 Dopo
201	200
231	236
221	216
260	233
228	224
237	216
326	296
235	195
240	207
267	247
284	210
201	209

CALCOLATE
LE DIFFERENZE
TRA LE DUE QUANTITA’

CALCOLATE
LA MEDIA E LA VARIANZA
DELLE DIFFERENZE

ADESSO POTETE ANDARE AVANTI

VERIFICA DI IPOTESI:
CONFRONTO DI DUE MEDIE PER CAMPIONI NON INDIPENDENTI



Campione

n=12
$\bar{d}$=-20,17
S^2_d=535,06

Hai ottenuto gli stessi risultati?

Se la risposta è NO ripeti i calcoli

Se la risposta è SI indica le……

Assunzioni:

☞ ……………..

☞ ……………

X_1	X_2	$d_i = x_2 - x_1$
201	200	-1
231	236	+5
221	216	-5
260	233	-27
228	224	-4
237	216	-21
326	296	-30
235	195	-40
240	207	-33
267	247	-20
284	210	-74
201	209	+8

$$\bar{d} = \frac{\sum d_i}{n}$$

$$S_d^2 = \frac{\sum \left(d_i - \bar{d}\right)^2}{n-1} = \frac{\sum d_i^2 - \left(\sum d_i\right)^2 / n}{n-1}$$

Dati: valori di colesterolo, in mg%ml, prima e dopo un trattamento di 12 soggetti. Si esegue la differenza tra i valori prima e dopo del trattamento.
La media è la media delle differenze,
la deviazione standard è la deviazione standard delle differenze.

Assunzioni:

☞ campione estratto da popolazione con distribuzione di Gauss;

☞ campioni non indipendenti

Ipotesi

H_0: $\mu_d = 0$

H_1: $\mu_d \neq 0$ → Ipotesi alternativa bidirezionale

$\mu_d > 0$ → Ipotesi alternativa unidirezionale

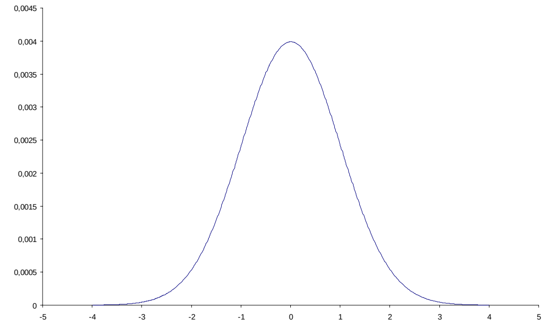
$\mu_d < 0$

Statistica test

$$t = \frac{\bar{d} - \mu_d}{S_d / \sqrt{n}}$$

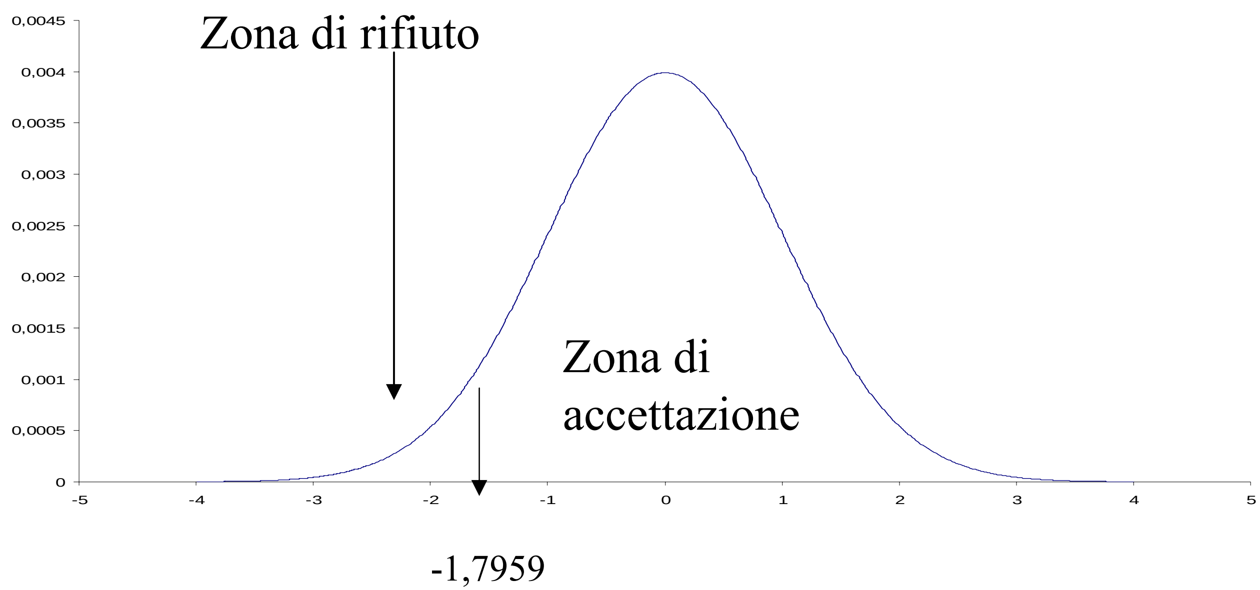
Distribuzione della statistica test

Distribuzione t-Student, caratterizzata dai gradi di libertà, n-1, della varianza



Regola di decisione

Fisso α=0,05, la probabilità di rifiutare l'ipotesi nulla quando è vera, Supponendo di aver formulato un ipotesi unidirezionale, si individua in questo modo il limite della zona di rifiuto (coda); sapendo che la distribuzione della statistica è t-student, il limite si ricerca nelle apposite tavole in corrispondenza del livello di significatività e dei gradi di libertà (n-1).



CALCOLATE
LA STATISTICA TEST

Poi verificate il risultato

Calcolo della statistica test:

$$t_{calc} = \frac{-20{,}17}{\sqrt{\frac{535{,}06}{12}}} = \frac{-20{,}17}{6{,}68} = -3{,}02$$

Decisione statistica

Rifiuto H_0 dato che $-3{,}02 < -1{,}7959$; il valore calcolato cade nella regione di rifiuto

Decisione clinica

Il trattamento eseguito ha avuto un effetto positivo nella riduzione dei livelli di colesterolo.

Si noti che se avessimo assunto l'indipendenza dei campioni sarebbero cambiati i gradi di libertà:

gruppo 1=12

gruppo 2=12

g.l.=$n_1+n_2-2=n+n-2=2n-2=2(n-1)=24-2=22$

Invece considerando i campioni appaiati si ha:

n=12 (numero delle differenze)

g.l.=n-1=12-1=11.

E' evidente che cambiando i gradi di libertà cambia anche il t_{tab}, con possibili conseguenze sull'accettazione o rifiuto dell'ipotesi nulla

Inoltre data la simmetria della distribuzione t-Student è indifferente effettuare le differenze tra le misure

Prima-Dopo oppure Dopo-Prima

SE NON FOSSE STATO POSSIBILE IPOTIZZARE

PER

LA VARIABILE “DIFFERENZA

TRA LE MISURE DEL COLESTEROLO”

UNA DISTRIBUZIONE DI GAUSS

COME AVRESTE DOVUTO EFFETTUARE

LA VERIFICA DELLE IPOTESI?

Riprendi i dati dell’esercizio e rispondi alla domanda prima di proseguire….

CONFRONTO TRA CAMPIONI NON INDIPENDENTI
TEST DELLA SOMMA DEI RANGHI CON SEGNO
(WILCOXON CON SEGNO)

X_1	X_2	$d_i=x_2-x_1$
201	200	-1
231	236	+5
221	216	-5
260	233	-27
228	224	-4
237	216	-21
326	296	-30
235	195	-40
240	207	-33
267	247	-20
284	210	-74
201	209	+8

Differenze senza segno in ordine non decrescente e relativo rango

1	1	-
4	2	-
5	3,5	+
5	3,5	-
8	5	+
20	6	-
21	7	-
27	8	-
30	9	-
33	10	-
40	11	-
74	12	-

Sommo tra loro i ranghi
con segno positivo **ΣR+= 8.5**
e quelli con segno negativo **ΣR- = 69.5**

Per prendere la decisione confronterò i valori calcolati con un intervallo di valori che trovo sulle tavole dei ranghi con segno in corrispondenza del numero di coppie e del livello di significatività $\alpha = 0.05$ stabilito a priori.

nel nostro esempio 13 – 65

Se la coppia di valori calcolati cade all'interno dell'intervallo tabulato allora accetto H_0;
se invece i valori cadono all'esterno dell'intervallo tabulato rifiuterò H_0.

Nel nostro caso entrambe le somme dei ranghi sono fuori dell'intervallo

pertanto

rifiuto l'ipotesi nulla e concludo che la dieta è stata efficace

OSSERVAZIONE IMPORTANTE!!!!!!

SE NEI CALCOLI RISULTASSERO
DIFFERENZE UGUALI A ZERO
QUESTE VANNO ELIMINATE,
PERCHE' ALLO ZERO NON E' POSSIBILE
ASSEGNARE UN SEGNO
E CONSEGUENTEMENTE SI RIDUCE LA
DIMESIONE DEL CAMPIONE

Si noti che a valori uguali si assegna rango pari alla media dei ranghi che ciascun valore avrebbe se fossero diversi

diff.		ranghi	segno delle diff.
1	**1**	**1**	**-**
4	**2**	**2**	**-**
5	**3** →	**3,5**	+
5	**4** →	**3,5**	**-**
8	**5**	**5**	+
20	**6**	**6**	**-**
21	**7**	**7**	**-**
27	**8**	**8**	**-**
30	**9**	**9**	**-**
33	**10**	**10**	**-**
40	**11**	**11**	**-**
74	**12**	**12**	**-**