STATISTICA MEDICA

Dott.ssa Marta Di Nicola
N.P.D. 3° Blocco 2° piano
0871-3554007
m.dinicola@unich.it
http://www.biostatistica.unich.it

I TEST D'IPOTESI

Statistica inferenziale per variabili quantitative

	Glicemia (mg/100cc)
x_1	103
X_2	97
X_3	90
x_4	119
X_5	107
X_6	71
X_7	94
X_8	81
X_9	92
x_{10}	96

Un problema pratico

Quesito: qual è la glicemia media della popolazione degli studenti iscritti al primo anno del corso di laurea in medicina?

- Procedura
  1. dall'elenco dei 200 studenti estraggo un campione casualmente costituito da 10 unità
  2. determino i livelli di glicemia in ciascun soggetto che costituisce il campioni
  3. calcolo la media aritmetica e la deviazione standard

Un problema pratico

Numerosità	n=10	N
Media aritmetica	$\bar{x}$=95	μ
Deviazione standard	s=13.32	σ

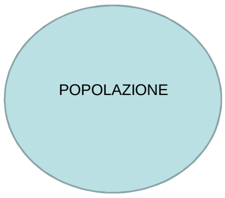
Posso affermare che = $\overline{X}$ = μ ?

NO!

perché l'estrazione del campione comporta necessariamente una perdita di informazione che si traduce nell'introduzione di un **errore casuale.**

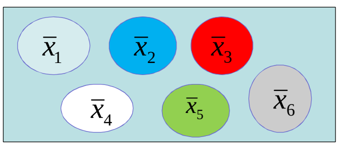
Un breve salto nella teoria

- La popolazione bersaglio è costituita da N elementi
- Il carattere in quantitativo in esame ha media aritmetica uguale a μ e deviazione standard uguale a σ



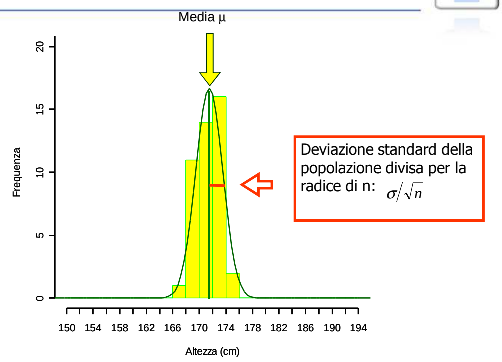
Un breve salto nella teoria

1. Si estraggano da tale popolazione infiniti campioni di numerosità n e per ognuno di essi si calcoli la media aritmetica;



2. il risultato è un insieme infinito di medie aritmetiche riferite a campioni di numerosità n;
3. se ciascuna media viene considerata come un'osservazione individuale è possibile costruire una distribuzione di frequenza delle medie campionarie.

Un breve salto nella teoria



La Distribuzione delle Medie Campionarie

Proprietà

1. La media della distribuzione delle medie campionarie è uguale alla media μ della popolazione;
2. La deviazione standard della distribuzione delle medie campionarie è uguale a $\sigma/\sqrt{n}$. Questa quantità è nota come *Errore Standard;*
3. La forma della distribuzione delle medie campionarie è approssimativamente normale, posto che *n* sia sufficientemente grande (n>30).

Formalizziamo il problema avanzando due ipotesi

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

Distribuzione di campionamento

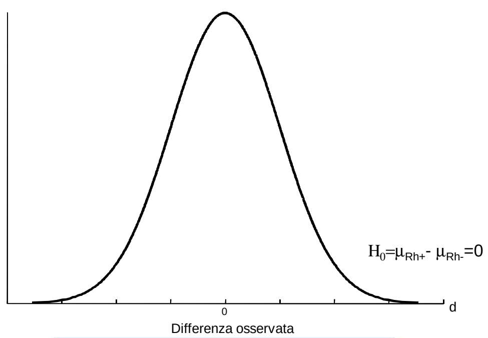
Distribuzione delle differenze campionarie teoricamente possibili se i gruppi fossero uguali.

1. È approssimativamente Gaussiana
2. La media della distribuzione è 0 ← Definita da H_0
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

Distribuzione di campionamento sotto H0



Il Test Statistico d'Ipotesi

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

Test statistico

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

Sotto l'ipotesi H_0 la differenza è nulla

La deviazione standard potremmo non conoscerla, ma sappiamo come stimarla

Il Test Statistico d'Ipotesi

Test statistico

Nel caso σ non sia nota possiamo stimarla con la S pooled:

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

$$s_{pooled} = \sqrt{\frac{s_1^2(n_1 - 1) + s_2^2(n_2 - 1)}{n_1 + n_2 - 2}}$$

Il Test Statistico d'Ipotesi

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

Si rifiuta H_0	Non si rifiuta H_0	Si rifiuta H_0

δ=0

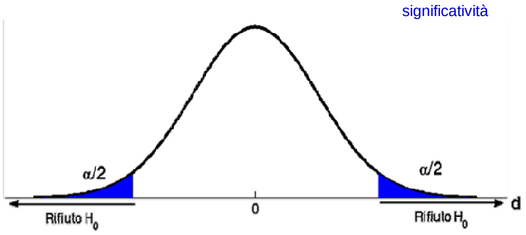
d

Distribuzione di campionamento sotto H0

Regione critica di rifiuto (2 code)

$H_0 = \mu_{Rh+} - \mu_{Rh-} = 0$

α = livello di significatività



Il Test Statistico d'Ipotesi

Rifiuto/non rifiuto di H_0

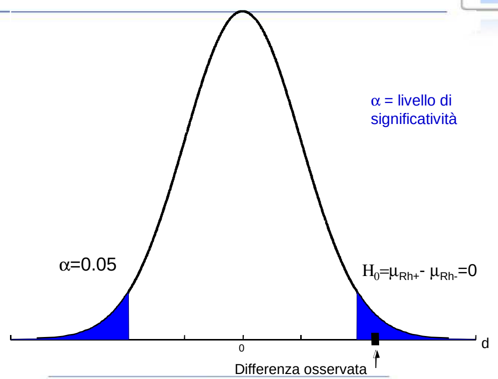
Con il Test Statistico calcoliamo *la probabilità* che la differenza osservata sia imputabile al caso

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 sia falsa.

Errore associato: risultato falso positivo o di I tipo

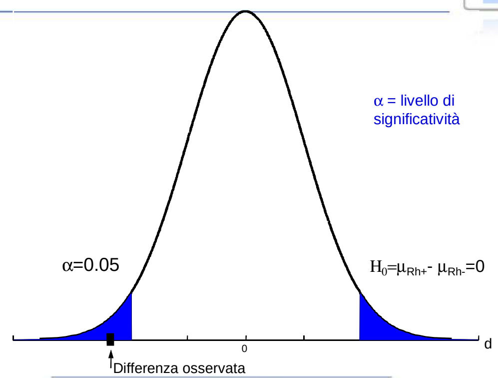
Risultato statisticamente significativo

α = livello di significatività



Risultato statisticamente significativo

α = livello di significatività



Test d'Ipotesi e p-value

Il risultato di un test statistico d'ipotesi può essere espresso in termini di rifiuto o non rifiuto di H_0

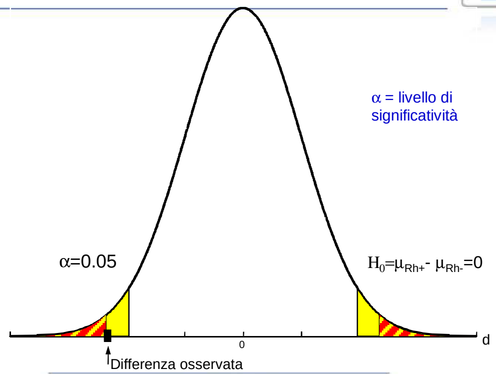
Statisticamente Significativo

Statisticamente Non Significativo

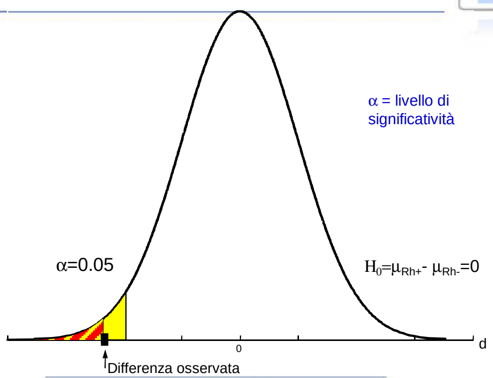
Oppure in termini di probabilità *p* che il valore osservato sia uguale o maggiore sotto l'ipotesi nulla

$P(X \leq -x \cup X \geq x \mid H_0) = p$

Risultato statisticamente significativo



Risultato statisticamente significativo



Il Test Statistico d'Ipotesi

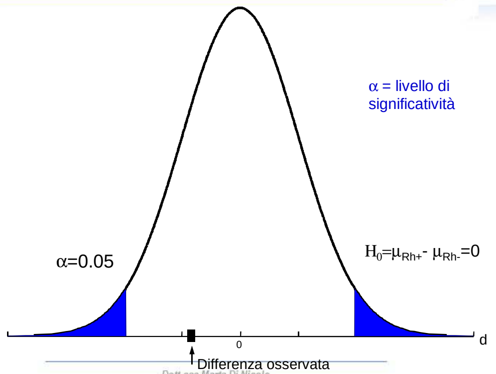
Rifiuto/non rifiuto di H_0

Con il Test Statistico calcoliamo *la probabilità* che la differenza osservata sia imputabile al caso

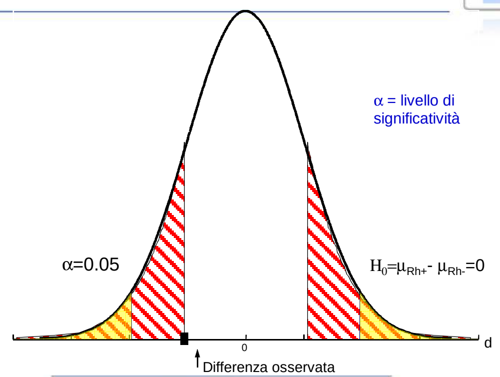
Se questa probabilità non è piccola, ovvero se il risultato osservato non è sufficientemente diverso da zero, il risultato si dice statisticamente non significativo: non abbiamo, cioè, prove sufficienti per confutare l'ipotesi nulla

Errore associato: risultato falso negativo o di II tipo

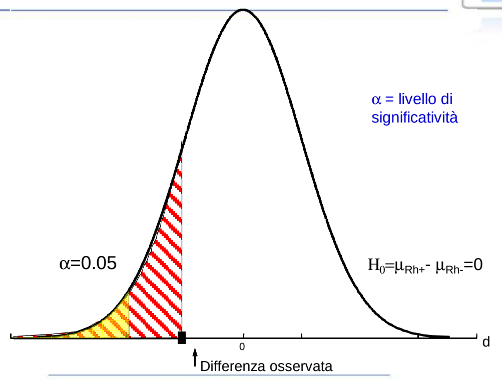
Risultato statisticamente non significativo



Risultato statisticamente non significativo



Risultato statisticamente non significativo



Il Test Statistico d'Ipotesi

Esempio

Il campione di studenti ha dato i seguenti risultati:

$n_+ = 15$
$n_- = 5$
$\bar{x}_+ = 171.3$
$\bar{x}_- = 166.6$
$S_{pooled} = 9.3$

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

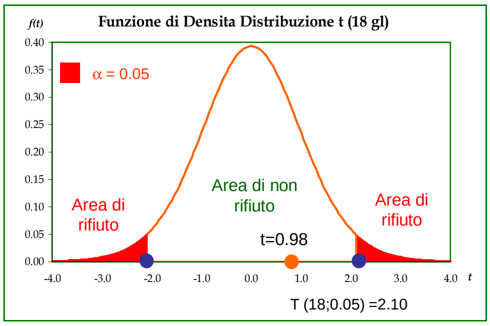
Esempio

Applichiamo il Test ai nostri dati osservati:

$$t=\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{s_{pooled}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \Rightarrow t=\frac{(171.3-166.6)-(0)}{9.3\cdot\sqrt{\frac{1}{15}+\frac{1}{5}}}=\frac{4.7}{4.8}=0.98$$

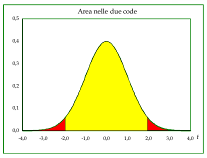
Il Test Statistico d'Ipotesi

Esempio



Percentili distribuzione t di Student

	PROBABILITA' (2 code)				PROBABILITA' (1 coda)			
GL	0,1	0,05	0,02	0,01	0,05	0,025	0,01	0,00 5
1	6,31	12,71	31,82	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

Campioni dipendenti

Formalizziamo il problema avanzando due ipotesi:

Ipotesi nulla (H_0)

Tecnicamente il confronto è semplice: l'analisi è ridotta alla sola serie risultante dalle differenze tra gli elementi di ciascuna coppia.

Se l'obiettivo è riconoscere un'eventuale differenza tra i gruppi, l'ipotesi nulla è che la media della popolazione delle differenze è 0

H_0: δ = 0 *ipotesi nulla*

H_1: δ ≠ 0 *ipotesi alternativa*

Il Test Statistico d'Ipotesi

Test statistico

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}}$$

Sotto l'ipotesi H_0 la differenza media (δ) è nulla

s_d è la deviazione standard delle differenze

Il Test Statistico d'Ipotesi

Esempio

Un gruppo di 10 cavie è stato sottoposto ad una dieta. Ogni soggetto è stato pesato prima e dopo la dieta:

Cavia	Peso prima	Peso dopo	Differenza (d)	(d-dmedio)2
1	180	190	10	1
2	175	170	-5	196
3	150	175	25	256
4	158	164	6	9
5	174	183	9	0
6	187	184	-3	144
7	172	185	13	16
8	157	168	11	4
9	164	180	16	49
10	165	173	8	1

La media delle differenza è pari a 9

Quanto è probabile che questa differenza sia imputabile al caso (se in realtà la dieta non ha effetto)?

Il Test Statistico d'Ipotesi

Esempio

Applichiamo il Test ai nostri dati osservati:

$$t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}} \Rightarrow t = \frac{9 - 0}{8.67 / \sqrt{10}} = 3.28$$

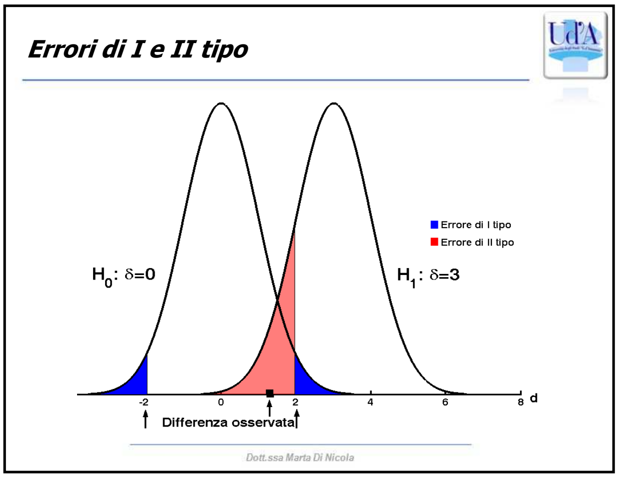
Il valore critico della distribuzione per 9 gradi di libertà e $\alpha = 0.05$ è pari a 2.262 (test a due code).

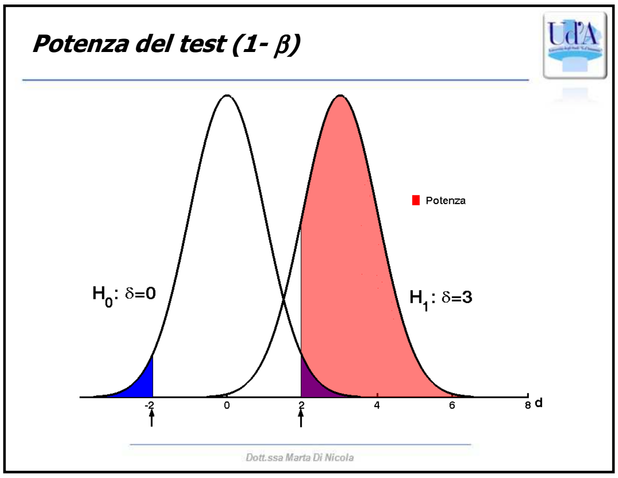
Dato che il valore di t è > del t tabulato si rifiuta H_0 : dunque la nuova dieta determina una differenza ponderale nelle cavie.

Errori di I e II tipo

Medie della Popolazione

Conclusione Test	Uguali (H_0)	Differenti (H_1)
Le medie *non sono* differenti	VN	FN II (β)
Le medie *sono* differenti	FP I (α)	VP potenza ($1-\beta$)


Errori di I e II tipo
Errore di I tipo
Errore di II tipo
H0: δ=0
H1: δ=3
-2
0
2
4
6
8
d
Differenza osservata
Dott.ssa Marta Di Nicola


Potenza del test (1- β)
Potenza
H0: δ=0
H1: δ=3
-2
0
2
4
6
8
d
Dott.ssa Marta Di Nicola