

Confronto di due misure

Campioni non indipendenti

o meglio

osservazioni “entro coppia”

**Es. il classico caso della farmacologia
nella quale si “misura” un animale
prima e dopo un trattamento.**

**O, meglio, si utilizza il soggetto (animale) come
controllo di se stesso!**

**in pratica non abbiamo
più due serie di misure
ma una sola serie
costituita da
differenze di misura**

quindi

differenze fra osservazioni entro la coppia!

$$x_1 - y_1$$

Le differenze seguiranno la distribuzione normale oscillando simmetricamente intorno allo \emptyset

Misura effettuata su “campione” x che segue la distribuzione normale

Misura effettuata su “campione” y che segue la distribuzione normale

Ipotesi zero \emptyset **cioè**

confronto analogo a “fra media campionaria e media della popolazione” nel caso particolare che la media della popolazione sia \emptyset .

Ipotesi zero

$t =$

$$d = x_1 - x_2$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})_A^2 + \sum (x - \bar{x})_B^2}{(n_A - 1) + (n_B - 1)}}$$

\bar{x} = media differenze fra dati appaiati $\frac{\sum (x_i - y_i)}{n}$

μ = zero

$$t = \frac{(\bar{x} - \mu)}{sm}$$

t di Student = $\frac{\text{media campione} - \text{media popolazione}}{\text{errore standard campione}}$

$$sm = \mathbf{d.s. / \sqrt{n}}$$

N.B. In questo caso n = n delle differenze cioè delle coppie di dati!

$$\sqrt{\frac{\sum [(x_i - y_i) - (\bar{x} - \bar{y})]^2}{(n - 1)}}$$

“Somma dei quadrati degli scarti delle differenze dalla media delle differenze”

analogamente

g.l. = n-1
delle coppie

$$\sqrt{\frac{\sum [(x_i - y_i) - (\bar{x} - \bar{y})]^2}{(n-1)}} = \sqrt{\frac{\sum (x_i - y_i)^2 - \frac{(\sum (x - y))^2}{n}}{(n-1)}}$$

cioè

La somma dei quadrati degli scarti delle differenze si può ottenere anche sottraendo alla somma dei quadrati delle differenze il “termine di correzione” cioè il quadrato della somma delle differenze diviso il numero dei dati della serie

Il termine: delle differenze ha semplicemente sostituito il termine dei dati nel modo di calcolo della d.s.

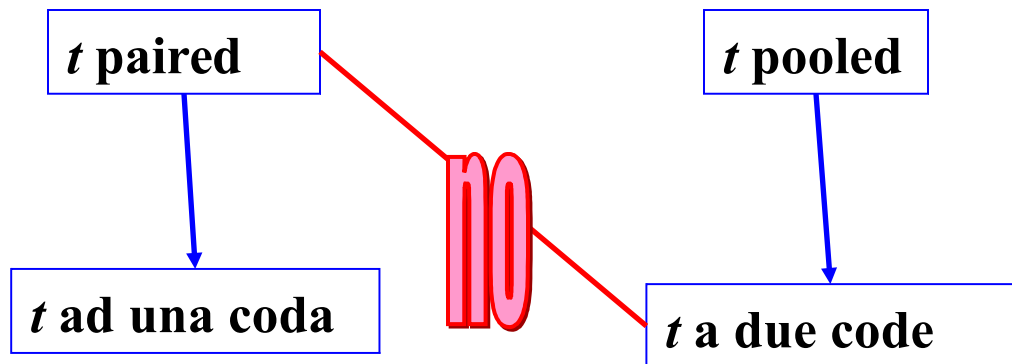
ricordiamo

$$\sum (\bar{x} - x)^2 = \sum (x^2 - C) \text{ dove } C = \frac{\sum x^2}{n}$$

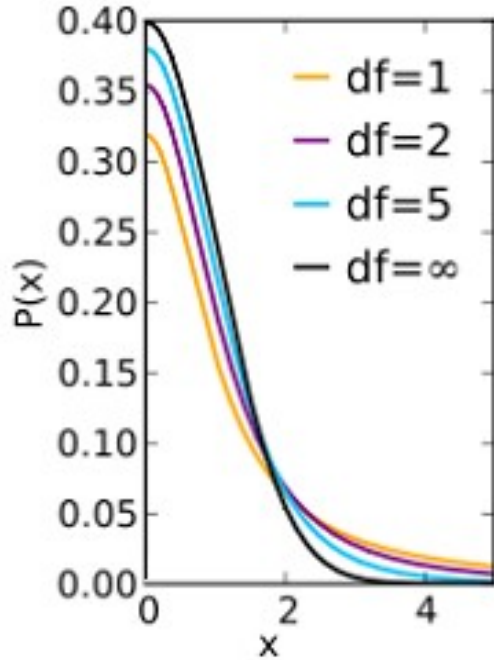
$$t = \frac{\frac{\sum (x_i - y_i)}{n}}{\sqrt{\frac{\sum (x_i - y_i)^2 - \frac{(\sum (x - y))^2}{n}}{n-1}}}$$

ma quale t ?

dai valori alle
differenze si "perde"
una coda



In questo caso:



$$\bar{X} + 0$$

$$\bar{X} - 0$$

	probabilità % di un valore più elevato di t trascurando il segno.									
due code	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,002	0,001
una coda	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
g.l.										
1	1,000	1,376	1,963	3,078	6,314	12,710	31,820	63,660	318,310	636,620
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626	3,174	3,390
1.000	0,675	0,842	1,037	1,282	1,646	1,962	2,330	2,581	3,098	3,300
infinito	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Esempio numerico

nome		Glicemia prima	Glicemia dopo
Ambra	n	70	110
Kira	n	85	80
Zoe	n	84	128
Telma	n	88	135
Flora	n	68	122
Pandora	n	77	100
Pizza	n	85	200
Birra	n	90	95
Turan	dia	240	200
Bianca	dia	120	160
Cecca	dia	130	130
Lolli	dia	226	240
Gigia	dia	135	185
Astra	dia	160	210
Bruna	dia	170	220
Kitti	dia	200	190
	n =	16	16

1

Calcolo i parametri del campione ed i parametri della differenze per coppie.

g.l. = $n-1$
delle coppie

$(n. \text{ dati}/2)-1$

2

Individuo il valore di t ad una coda per $(32/2)-1=15$ g.l.

probabilità % di un valore più elevato di t trascurando il segno.										
due code	0,5	0,4	0,3	0,2	0,1	0,05	0,02	0,01	0,002	0,001
una coda	0,25	0,2	0,15	0,1	0,05	0,025	0,01	0,005	0,001	0,0005
g.l.										
1	1,000	1,376	1,963	3,078	6,314	12,710	31,820	63,660	318,310	636,620
2	0,816	1,061	1,386	1,886	2,920	4,303	6,965	9,925	22,327	31,599
3	0,765	0,978	1,250	1,638	2,353	3,182	4,541	5,841	10,215	12,924
4	0,741	0,941	1,190	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	0,920	1,156	1,476	2,015	2,571	3,355	4,032	5,893	6,869
6	0,718	0,906	1,134	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	0,896	1,119	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	0,889	1,108	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	0,883	1,100	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	0,879	1,093	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	0,876	1,088	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	0,873	1,083	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	0,870	1,079	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	0,868	1,076	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	0,866	1,074	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	0,865	1,071	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	0,863	1,069	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	0,862	1,067	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	0,861	1,066	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	0,860	1,064	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	0,859	1,063	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	0,858	1,061	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	0,858	1,060	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	0,685	0,857	1,059	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	0,856	1,058	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	0,856	1,058	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	0,855	1,057	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	0,855	1,056	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	0,854	1,055	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	0,854	1,055	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	0,851	1,050	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	0,848	1,045	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,678	0,846	1,043	1,292	1,664	1,990	2,374	2,639	3,195	3,416
100	0,677	0,845	1,042	1,290	1,660	1,984	2,364	2,626	3,174	3,390
1.000	0,675	0,842	1,037	1,282	1,646	1,962	2,330	2,581	3,098	3,300
infinito	0,674	0,842	1,036	1,282	1,645	1,960	2,326	2,576	3,090	3,291

3

Calcolo: $t = \frac{\text{differenza media da zero}}{\text{errore standard campione}}$

	Glicemia prima	Glicemia dopo	diff
n =	16	16	16
media =	126,75	156,5625	29,8125
SS=	48.455,00	37.753,94	19.240,44
VAR =			1.282,70
d.s. =	56,836021442	50,169006	35,8147432
es =		es =	8,95368581
$n_A + n_B - 2 =$	30	= g.l.	15
calcolato =	1,5729969153	t calcolato =	3,32963437
$t_{0,05} =$	2,042	$t_{0,05} =$	1,753
$t_{0,01} =$	2,750	$t_{0,01} =$	2,602

$$sm = d.s./\sqrt{n}$$

4

Confronto t calcolato con t da tabella per sapere se è compreso o meno nella oscillazione casuale

Se calcolassi il t pooled (In questo caso sarebbe **SBAGLIATO**) infatti verrebbe non significativo!

La glicemia “prima” differisce in modo altamente significativo da quella “dopo”.

Osserva come in questo caso diventerebbe “difficile” calcolare la somma dei quadrati degli scarti tramite la definizione invece che tramite il T.C.!

		Glicemia prima	Glicemia dopo	diff	quadrati	scarti	quadrati scarti
Ambra	n	70	110	40	1600	10,1875	103,7851563
Kira	n	85	80	-5	25	-34,8125	1211,910156
Zoe	n	84	128	44	1936	14,1875	201,2851563
Telma	n	88	135	47	2209	17,1875	295,4101563
Flora	n	68	122	54	2916	24,1875	585,0351563
Pandora	n	77	100	23	529	-6,8125	46,41015625
Pizza	n	85	200	115	13225	85,1875	7256,910156
Birra	n	90	95	5	25	-24,8125	615,6601563
Turan	dia	240	200	-40	1600	-69,8125	4873,785156
Bianca	dia	120	160	40	1600	10,1875	103,7851563
Cecca	dia	130	130	0	0	-29,8125	888,7851563
Lolli	dia	226	240	14	196	-15,8125	250,0351563
Gigia	dia	135	185	50	2500	20,1875	407,5351563
Astra	dia	160	210	50	2500	20,1875	407,5351563
Bruna	dia	170	220	50	2500	20,1875	407,5351563
Kitti	dia	200	190	-10	100	-39,8125	1585,035156
		n =	16	16		16	16
		somme=	2505	477	33461	0	
		quadrato		227529			
		media =	126,75	156,5625			
con "TC"	con "TC" SS=			19240,4375			
	con definizione SS=	48.455,00	37.753,94	19.240,44			19.240,44
	VAR =			1.282,70			1.282,70
	d.s. =	56,83602144	50,16900604	35,81474324			35,81474324
				8,95368581			8,95368581

	T Randomizzato	t appaiato
d (A-B)=	29,8125	
$n_A + n_B =$	32	
$n_A * n_B =$	256	
$n_A + n_B - 2 =$	30	15
	3,75	
$ds_d^2 =$	359,2039063	
$ds_d =$	18,95267544	
t calcolato =	1,572996915	3,329634369
$t_{0,05} =$	2,042	$t_{0,05} = 1,753$
$t_{0,01} =$	2,750	$t_{0,01} = 2,602$
P =	0,120207986	P = 0,002286207
$MDS_{0,05} =$	38,70130324	
$MDS_{0,01} =$	52,11985745	

5

Riporto i risultati

Prima cifra significativa dell'e.s. sulle unità

T-paired test	Glicemia prima		Glicemia dopo	
n =	16		16	
media =	126,8	A	156,6	B
d.s. =	56,84		50,17	
nota: lettere diverse indicano differenze significative per $p < 0,05$ se minuscole; $p < 0,01$ se maiuscole.				

Oppure.....

MEGLIO e per di più più facile!

5

Riporto i risultati

T-paired test	Glicemia prima	Glicemia dopo	differenza	
n =	16	16	16	
media =	126,8	156,6	29,8	**
d.s. =			35,81	
nota: * valore significativamente diverso da zero per $p < 0,05$ ** valore significativamente diverso da zero per $p < 0,01$.				

Riassumendo

t appaiato (calcolo più semplice) poiché si deve vedere se c'è differenza entro lo stesso animale trattato, basta fare:

- la differenza per ogni animale dei due valori accoppiati della serie
- calcolare l'errore standard della serie di differenze così calcolate
- calcolare la media della serie di differenze così calcolate
- dividere la media per l'errore standard
- confrontare il t calcolato con quello tabulato **AD UNA CODA**

t randomizzato (calcolo più complesso) da utilizzare sempre a meno che le misure non siano “accoppiate” (fatte sullo stesso animale) è necessario:

- calcolare la somma dei quadrati degli scarti della prima e seconda serie di misure
- sommare le due somme e poi dividerle per la somma dei due gradi di libertà quindi calcolarne la radice quadrata (in sostanza si tratta di calcolare correttamente la "media" delle due deviazioni standard)
- calcolare il rapporto fra la somma delle due numerosità ed il prodotto delle due numerosità e moltiplicare la radice quadrata di tale rapporto per la "media" delle due deviazioni standard calcolata precedentemente (in sostanza si tratta di calcolare correttamente la "media" dei due errori standard)
- calcolare la differenza fra le medie
- dividere la differenza fra le medie per la "media" dei due errori standard
- confrontare il valore trovato “ t calcolato” con quello tabulato **A DUE CODE**

