

STATISTICA MEDICA 

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

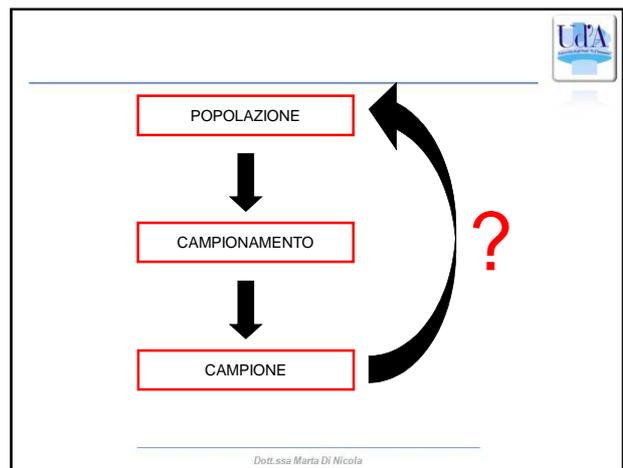
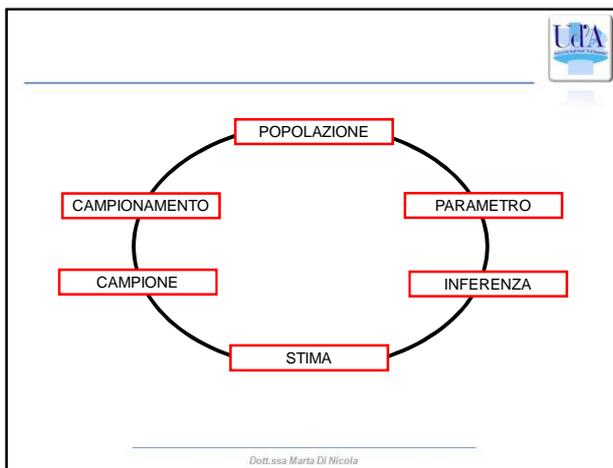
<http://www.biostatistica.unich.it>

Dott.ssa Marta Di Nicola



**L'INTERVALLO DI
CONFIDENZA**

Dott.ssa Marta Di Nicola



Stimare i Parametri della Popolazione



- La media del gruppo (campione) è una *stima puntuale* del parametro della popolazione
- Ogni media di gruppo fornisce una diversa stima connessa alle *fluttuazioni casuali* dovute al campionamento
- La stima puntuale non da indicazioni sulla variabilità della stima
- Costruisco un intervallo centrato intorno alla media di gruppo sul quale ho una certa *confidenza* che il parametro della popolazione cada nell'intervallo
- L'intervallo di confidenza è la *stima intervallare* del parametro della popolazione

Dott.ssa Marta Di Nicola

Intervallo di Confidenza e Parametro



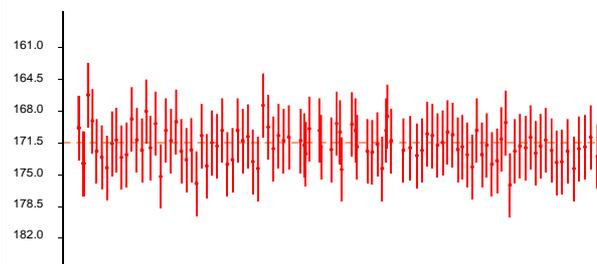
Dott.ssa Marta Di Nicola

Intervallo di Confidenza e Parametro



Dott.ssa Marta Di Nicola

Intervallo di Confidenza e Parametro



Dott.ssa Marta Di Nicola

Intervallo di Confidenza

- Gli intervalli di confidenza sono definiti come un intervallo di valori costruito a partire dai dati
- All'interno dell'intervallo ho una certa probabilità (tipicamente 95%) che sia compreso il parametro della popolazione

Dott.ssa Marta Di Nicola

Intervallo di Confidenza e Parametro

Nel 95% circa dei campioni possibili l'intervallo di confidenza al 95% comprende il parametro della popolazione (171.5 cm)

Dott.ssa Marta Di Nicola

Intervallo di Confidenza

- Gli intervalli di confidenza sono definiti come un intervallo di valori costruito a partire dai dati
- All'interno dell'intervallo ho una certa probabilità (tipicamente 95%) che sia compreso il parametro della popolazione
- Per calcolare l'intervallo utilizzo le proprietà della distribuzione di campionamento delle medie

$$\Pr\left\{\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

Dott.ssa Marta Di Nicola

Esempio di Calcolo dell'Intervallo di Confidenza al 95%

$$\Pr\left\{\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

Informazioni

$n = 20$	Limite Inferiore
$\bar{x} = 170$	$170 - 1.96 \cdot 8.5 / \sqrt{20} = 170 - 3.72 = 166.28$
$\sigma = 8.5$	Limite Superiore
$z = 1.96$	$170 + 1.96 \cdot 8.5 / \sqrt{20} = 170 + 3.72 = 173.72$

Dott.ssa Marta Di Nicola

Intervallo di Confidenza

Proprietà

- Maggiore è l'ampiezza dell' Intervallo di Confidenza minore è la precisione della stima
- La sua ampiezza, e quindi la precisione della stima, varia con la numerosità dello studio e il grado di confidenza desiderato
 - All'aumentare della numerosità l'ampiezza diminuisce e la precisione aumenta
 - All'aumentare del grado di confidenza (es. 99% invece di 95%) l'ampiezza aumenta e la precisione diminuisce

Dott.ssa Marta Di Nicola

Se σ è sconosciuta ?

Problema

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

(NB se μ non è nota, è probabile che anche σ^2 non sia nota)

Soluzione

Utilizzo la varianza campionaria s^2 come stima di σ^2

(NB nella formula della varianza divido per $(n-1)$: i gradi di libertà)

Dott.ssa Marta Di Nicola

La distribuzione t di Student

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad \longrightarrow \quad t = \frac{\bar{X} - \mu}{s / \sqrt{n}}$$

Il nuovo rapporto standardizzato non ha una distribuzione normale standardizzata perché devo tener conto anche della variabilità di s che sarà maggiore quando n è piccolo. Questo rapporto è distribuito come una t di Student con n-1 gradi di libertà

Dott.ssa Marta Di Nicola

Percentili della distribuzione t di Student

GL	PROBABILITÀ (2 code)				PROBABILITÀ (1 coda)			
	0,1	0,05	0,02	0,01	0,05	0,025	0,01	0,005
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	2,28	2,82	3,25	1,83	2,28	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

Dott.ssa Marta Di Nicola

La distribuzione t di Student



Caratteristiche

1. È una distribuzione continua
2. È simmetrica rispetto alla media: μ
3. Media, mediana e moda coincidono
4. È una distribuzione di *probabilità*
5. Se n è basso i valori nelle code sono più probabili
6. Al crescere di n la distribuzione approssima la *gaussiana standardizzata*

Dott.ssa Marta Di Nicola

Distribuzione t di Student e Intervallo di Confidenza



Consideriamo i dati sull'altezza raccolti da un gruppo di studenti

$$\begin{aligned} n &= 20 \\ \bar{x} &= 172.0 \\ s &= 10.0 \end{aligned}$$

Qual è l'intervallo di confidenza al 95% della media ?

Dott.ssa Marta Di Nicola

Distribuzione t di Student e Intervallo di Confidenza



Occorre modificare la formula precedente

$$\Pr\left\{\bar{X} - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right\} = 0.95$$

tenendo conto delle nuove informazioni

$$\Pr\left\{\bar{X} - t_{n-1} \cdot \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t_{n-1} \cdot \frac{S}{\sqrt{n}}\right\} = 0.95$$

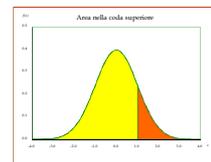
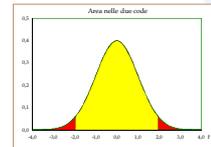
Quali valori della distribuzione t di Student con 19 gradi di libertà lasciano un'area nelle due code pari a 0.05 ?

Dott.ssa Marta Di Nicola

Percentili della distribuzione t di Student



GL	PROBABILITÀ (z code)					PROBABILITÀ (t code)				
	0.1	0.05	0.02	0.01	0.005	0.01	0.005	0.01	0.005	
1	6.31	12.71	19.22	63.66	6.31	12.71	19.22	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	1.79	2.18	2.68	3.05	1.79	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		



Dott.ssa Marta Di Nicola

Calcolo dell'Intervallo di Confidenza 

Inseriamo le informazioni raccolte nella formula

$$\left(\bar{X} - t_{n-1} \cdot \frac{s}{\sqrt{n}}, \bar{X} + t_{n-1} \cdot \frac{s}{\sqrt{n}} \right)$$
$$\left(172 - 2.09 \cdot \frac{10}{\sqrt{20}}, 172 + 2.09 \cdot \frac{10}{\sqrt{20}} \right)$$

I limiti dell'intervallo di confidenza sono 167.33 e 176.33

Dott.ssa Marta Di Nicola