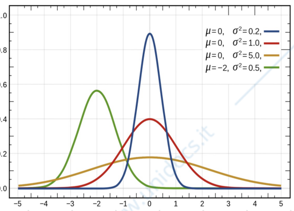
INTERVALLI DI FIDUCIA

• INCERTEZZA: la possiamo esprimere anche attraverso INTERVALLI DI FIDUCIA (modo più raffinato rispetto alla deviazione standard)
↓
Teniamo in considerazione la probabilità con cui si distribuiscono le nostre misurazioni replicate intorno alla media

Per definire gli intervalli di fiducia utilizziamo il concetto di DISTRIBUZIONE di PROBABILITÀ (in particolare, la distribuzione Gaussiana, la distribuzione normale standard standard e la distribuzione t student). Attraverso ciò associamo un grado di probabilità al nostro risultato

DISTRIBUZIONE NORMALE o GAUSSIANA



i due parametri fondamentali sono:

• μ : centro della distribuzione (MEDIA)

• σ : ampiezza distribuzione (DEVIAZIONE STANDARD)

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}$$

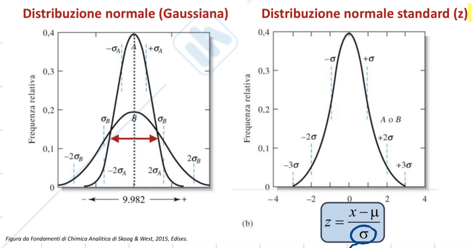
Funzione densità di probabilità

A seconda dell'incertezza che abbiamo le curve saranno più strette o più larghe

Noi standardizziamo le distribuzioni Gaussiane attraverso la DISTRIBUZIONE NORMALE STANDARD
↓
È una distribuzione gaussiana standard con una variabile z (Rappresenta ogni curva gaussiana)

$$z = \frac{x-\mu}{\sigma} \quad \Rightarrow f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-1/2\,(z)^2}$$

EX: a sx abbiamo 2 curve Gaussiane con stessa μ ma $\sigma \neq$



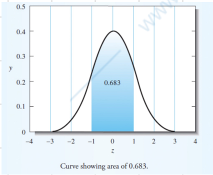
Attraverso la trasformazione z ottengo la distribuzione di dx. È sempre centrata sullo 0 (effetto dell'operazione $x-\mu$)

la divisione per σ ci permette di ottenere delle unità che corrispondono alle deviazioni standard della nostra curva

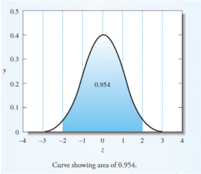
Grazie al termine $\frac{1}{\sqrt{2\pi}}$ sappiamo che l'area sottesa dalla curva è = 1

Cioè $F(x) = \int_{-\infty}^{+\infty} f(x)\,dx = 1$

Noi possiamo associare delle probabilità ai nostri valori di x o z, l'area sottesa a una curva tra due limiti corrisponde alla probabilità che un valore misurato sia compreso tra quei limiti.



Curve showing area of 0.683.



Curve showing area of 0.954.

Se restringo l'intervallo la probabilità scende nel caso a dx, dato che l'intervallo va da −2 a 2 ho il 95% di probabilità.

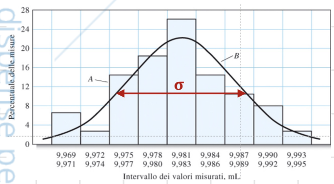
Una volta che standardizzo la mia grandezza so che al 95% di probabilità la mia misurazione replicata deve essere compresa tra questi due limiti, così associo alla mia incertezza un grado di probabilità definendo gli INTERVALLI di CONFIDENZA

Per una variabile standard z, la probabilità può essere calcolata utilizzando la tavola della distribuzione standard di probabilità.

Al variare di z so qual è l'area sottesa alla curva.

z	P % (fino a)	(1 - P) %	P % (compreso tra)
0.00	50.00	50.00	0.00
0.67	74.86	25.14	49.72
1.00	84.13	15.87	68.26
1.28	89.97	10.03	79.94
1.64	94.95	5.05	89.90
1.96	97.50	2.50	95.00
2.00	97.72	2.28	95.44
2.58	99.51	0.49	99.02
3.00	99.87	0.13	99.74
3.29	99.95	0.05	99.90

DEVIAZIONE STANDARD della POPOLAZIONE (σ): Variabilità di un numero altissimo di misurazioni replicate.
→ Noi raramente conosceremo σ



Noi facciamo poche misurazioni replicate, dalle quali troviamo una stima attraverso s (DEVIAZIONE STANDARD)

$$s = \sqrt{\frac{\sum_{i=1}^{m}(x_i-\bar{x})^2}{m-1}}$$

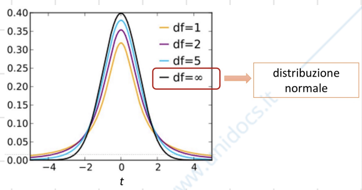
all'aumentare di "m", s diventa una migliore stima di σ ($s \rightarrow \sigma$)

Per un PICCOLO CAMPIONE o stime ottenute da POCHE RIPETIZIONI, si approssima la distribuzione normale standard con la DISTRIBUZIONE t di STUDENT

utilizziamo la trasformazione $\boxed{t = \frac{x-\mu}{s}}$

le distribuzioni t dipendono dai GRADI di LIBERTÀ (d.f.)

n° di misurazioni replicate − 1, con cui ho stimato la deviazione standard



per d.f.=1 (2 repliche) la curva è più larga e c'è più incertezza.

d.f.=2 la curva si stringe

! Mano a mano che aumento le repliche la curva si stringe (tutte le curve sottendono un'area di 1).

CASO LIMITE: d.f = ∞, aumentando notevolmente il n° di repliche otteniamo la distribuzione normale standard

La distribuzione t possiamo valutarla in 2 situazioni:

- SINGOLA MISURAZIONE: $t = \frac{x-\mu}{s}$
- SERIE di MISURAZIONI: $t = \frac{\bar{x}-\mu}{s_m} = \frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$ (utilizzo i valori medi al posto del singolo)

Determinare il valore vero (μ) della grandezza misurata richiederebbe troppe misurazioni, tuttavia possiamo stabilire un intervallo intorno alla media sperimentale ($\bar{x}$) all'interno del quale si prevede che si trovi μ. Tale intervallo è chiamato INTERVALLO di FIDUCIA.
↓
L'intervallo di confidenza della media è un'incertezza.
Il suo centro è la media e noi lo definiamo su una probabilità

EX: << misurazioni dei ppm di Fe su 10 misurazioni >>

troviamo $\bar{x} = 30$ ppm. Questa media sarà il centro dell'intervallo

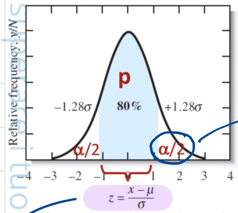
Per esempio potremmo avere un intervallo di confidenza del 95% con incertezza ±2.

Mi aspetto che il valore del Fe sia compreso tra (28 e 32) ppm al 95% di probabilità.

LIVELLO di CONFIDENZA: è la probabilità (p) che il valore vero (μ) della grandezza misurata sia incluso nell'intervallo tra i valori $-z_p$ e $+z_p$. Si esprime come $1-\alpha$
α è il LIVELLO di SIGNIFICATIVITÀ

$$\boxed{p = \int_{-z_p}^{+z_p} f(z)\,dz}$$

Se consideriamo un livello di fiducia al 90%, $\alpha = 10\%$
Se consideriamo un livello di fiducia al 85%, $\alpha = 15\%$
e così via...



→ Se definiamo un livello di fiducia dell'80% avremo $\alpha = 20\%$

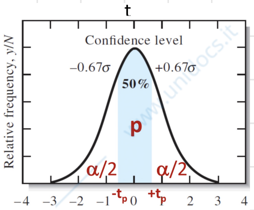
$\frac{\alpha}{2} = 10\%$

Avendo una curva standardizzata io so che all'80% il mio valore vero sarà compreso tra $-z_p = -1{,}28\sigma$ e $+z_p = +1{,}28\sigma$

Se non c'è ERRORE SISTEMATICO mi aspetto che μ sia al centro.
↓
definisco un intervallo, in funzione di una probabilità, che contenga il valore vero

Si basa su questo dato σ, ne occorre una stima ragionevole

▷ Consideriamo il caso in cui abbia un PICCOLO CAMPIONE o stime ottenute da POCHE RIPETIZIONI



• INTERVALLO di CONFIDENZA: è la PROBABILITÀ (p) che il valore vero (μ) della grandezza misurata sia incluso nell'intervallo tra i valori $-t_p$ e $+t_p$

Noi, in questo caso, non possiamo stimare σ ma conosciamo s (deviazione standard del campione rappresentato dalle misure replicate).

Quindi mi baserò sulla rappresentazione t-student: $\boxed{p=\int_{-t_p}^{+t_p} f(t)\,dt}$

Le misurazioni sono stimate attraverso s: $t=\frac{x-\mu}{s}$ (per una singola misurazione)

Per una serie di misurazioni replicate avremo $t=\frac{\bar{x}-\mu}{s_m}=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{m}}}$

Tuttavia la scelta di un determinato intervallo di fiducia non esclude totalmente la possibilità di fare previsioni sbagliate: se abbiamo scelto il livello di fiducia ($1-\alpha$) con $p=95\%$, avremo sempre il 5% di possibilità che il valore cada al di fuori dell'intervallo di fiducia.

Calcoliamo un intervallo di fiducia: $-z_p<z<z_p$, se io ponessi $z=\frac{x-\mu}{\sigma}$ conoscendo σ e x

otterrei $-z_p<\frac{x-\mu}{\sigma}<+z_p$ → $\underbrace{x-z_p\cdot\sigma}_{\text{limite inferiore}}<\mu<\underbrace{x+z_p\cdot\sigma}_{\text{limite superiore}}$ → intervallo centrato in x con una semiampiezza pari a $z_p\cdot\sigma$

▷ Per una singola misurazione: $x-z_p\sigma<\mu<x+z_p\sigma$

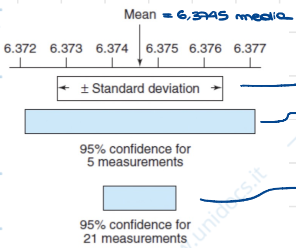
▷ Media di "m" misure: $\bar{x}-z_p\cdot\frac{\sigma}{\sqrt{m}}<\mu<\bar{x}+z_p\cdot\frac{\sigma}{\sqrt{m}}$

Mentre se σ non fosse nota, ma stimata con s, avremo un intervallo di confidenza:

▷ Per una singola misurazione: $x-t_p\cdot s<\mu<x+t_p\cdot s$

▷ Per la media di "m" misure: $\bar{x}-t_p\cdot\frac{s}{\sqrt{m}}<\mu<\bar{x}-t_p\cdot\frac{s}{\sqrt{m}}$

Intervallo di fiducia intorno alla media $\bar{x}$ ottenuta da m repliche



→ Deviazione standard della media

→ Intervallo di confidenza per 5 misurazioni, è una stima precauzionale poiché sto ampliando l'intervallo di incertezza legato alla misura

→ Aumentando il n° di repliche riduco la semiampiezza dell'intervallo di confidenza

In questo modo fornisco un intervallo più preciso.

Stima valore vero — Stima incertezza

$\bar{x} - t_p \cdot \frac{s}{\sqrt{m}} < \mu < \bar{x} - t_p \cdot \frac{s}{\sqrt{m}}$ → lo posso riscrivere: $\bar{X} = \pm t_{\alpha(2),v} \cdot S_m$ → S_m è la deviazione standard della media

media di x

valore della t-Student al livello di significatività $\alpha(2)$ → notazione statistica

V = gradi di libertà (m − 1)

Se ho un intervallo del 95% ognuna di queste due aree sarà 2,5%, che sommate danno il 5% → Indica che ci sono 2 porzioni bianche ai lati del nostro intervallo di fiducia

EX 1: « 25 misurazioni replicate della concentrazione di mercurio (ppb) di un campione di acqua fluviale forniscono le seguenti stime. Calcolare l'intervallo di confidenza della media con un livello di fiducia del 95% (significatività 5%) »

(si dovranno disporre con una distribuzione gaussiana intorno alla media)

Noi conosciamo $m = 5$; media = 24,23; varianza = 18,06 → $s = \sqrt{18,06} = 4,25$

Noi vogliamo stimare un intervallo di fiducia del 95% → $S_m = \frac{s}{\sqrt{m}} = \frac{4,25}{\sqrt{25}} = 0,85$

Ora devo trovare t, sapendo che i gradi di libertà sono 24, trovo il valore di t sulla tabella

TABLE B.3 CRITICAL VALUES OF THE t DISTRIBUTION

ν	α(2): 0.50 α(1): 0.25	0.20 0.10	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	0.005 0.0025	0.002 0.001	0.001 0.0005
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725

Devo conoscere:

- gradi di libertà → righe
- livello di confidenza → colonne (il livello di significatività α)

Quindi per un livello di fiducia del 95% (1−0,05), per 24 gradi di libertà avrò:

$t_{\alpha(2),v} = t_{0,05(2),24} = 2,064$, la semiampiezza dell'intervallo è data da $\bar{X} \pm t_{\alpha(2),v} \cdot S_m$

⇒ $24,23 \pm 2,064 \cdot 0,85 = 24,23 \pm 1,75$ abbiamo → • LIMITE SUP = 25,98
• LIMITE INF = 22,48

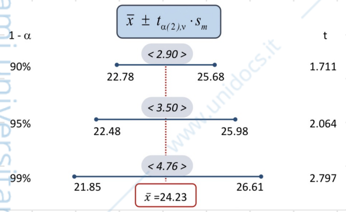
▷ Mentre se avessimo un intervallo di fiducia del 90% (1−0,10):

$t_{\alpha(2),v} = t_{0,10(2),24} = 1,711$ ⇒ LIMITE SUP: 25,68 e LIMITE INF = 22,78

▷ Per un livello di fiducia del 99% (1−0,01) avremo:

$t_{\alpha(2),v} = t_{0,01(2),24} = 2,797$ ⇒ LIMITE SUP: 21,85 e LIMITE INF = 26,61

t cresce al crescere del livello di significatività, aumentando il livello di fiducia l'intervallo aumenta



Noi stiamo aumentando la probabilità con cui vogliamo assumere che il valore vero che stiamo cercando sia all'interno del nostro intervallo.

EX 2: « Un clinico ha ottenuto i seguenti risultati relativi al contenuto di alcool (C_2H_5OH) in un campione di sangue:

0,084% 0,089% 0,079%

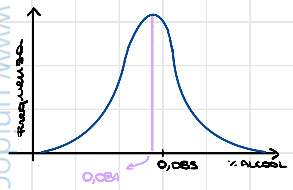
Calcolare l'intervallo di fiducia della media al 95% e al 99%, assumendo che i 3 risultati siano l'unico modo per stimare la precisione del metodo »

noi non conosciamo σ ma abbiamo stimarla attraverso s

Anche in questo caso useremo la distribuzione t di student per stimare il nostro intervallo di confidenza.

Abbiamo bisogno della media e di s:

$\bar{x} = \frac{\sum_{i=1}^{3} x_i}{3} = \frac{0{,}084 + 0{,}089 + 0{,}079}{3} = 0{,}084\%$ e $s = \sqrt{\frac{\sum_{i=1}^{3}(x_i - \bar{x})^2}{3-1}} = \sqrt{\frac{(0{,}084-0{,}084)^2 + (0{,}089-0{,}084)^2 + (0{,}079-0{,}084)^2}{2}} = 0{,}005\%$



Ora standardizzo questa distribuzione in una distribuzione t-student per poter ricavare i valori di t dalla tabella

Avendo 3 repliche abbiamo V=2 gradi di libertà

Per il 95% → $t_{0,05(2),2} = 4{,}303$ e per il 90% → $t_{0,10(2)2} = 9{,}925$

Intervalli corrispondenti a:

- 95%: $\bar{x} = \pm t_{0,05(2),2} \cdot \frac{s}{\sqrt{n}} = 0{,}084 \pm 4{,}303 \cdot \frac{0{,}05}{\sqrt{3}} = 0{,}084 \pm 0{,}012\%$
- 90%: $\bar{x} = \pm t_{0,10(2),2} \cdot \frac{s}{\sqrt{n}} = 0{,}084 \pm 9{,}925 \cdot \frac{0{,}05}{\sqrt{3}} = 0{,}084 \pm 0{,}029\%$

semiampiezze

noi li abbiamo stimati presupponendo che i 3 risultati fossero l'unico modo per conoscere l'incertezza legata ai risultati. Tuttavia in alcune situazioni noi possiamo conoscere σ, ad esempio se abbiamo un metodo che applichiamo sempre per serie molto alte di campioni possiamo avere una grande quantità di dati con cui possiamo presupporre di conoscere σ.

Per esempio sappiamo che $\boxed{\sigma = 0{,}005\%}$, in questo caso possiamo usare la distribuzione normale standard e stimare la media come: $\bar{x} = \pm z_p \cdot \frac{\sigma}{\sqrt{n}}$

Ci manca solo conoscere z_p (valori critici della distribuzione normale standard)

Tabella 5–2
Livello di confidenza per vari valori di z

Livello di confidenza %	z
50	0.67
68	1.00
80	1.29
90	1.64
95	1.96
96	2.00
99	2.58
99.7	3.00
99.9	3.29

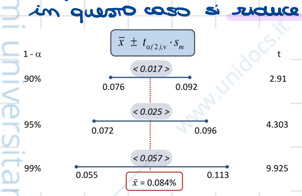
In questo caso non ci serve conoscere i gradi di libertà, poiché assumiamo che essi siano infiniti

Aumentando il livello di confidenza aumenta il valore critico.

Ora ricalcolo gli intervalli:

- 95%: $\bar{x} = \pm z_{95} \cdot \frac{\sigma}{\sqrt{n}} = 0{,}084 \pm 1{,}96 \cdot \frac{0{,}005}{\sqrt{3}} = 0{,}084 \pm 0{,}006\%$
- 99%: $\bar{x} = \pm z_{99} \cdot \frac{\sigma}{\sqrt{n}} = 0{,}084 \pm 2{,}58 \cdot \frac{0{,}005}{\sqrt{3}} = 0{,}084 \pm 0{,}013\%$

in questo caso si riduce l'incertezza associata al mio risultato, poiché i nostri intervalli si basano su una stima molto più affidabile dell'incertezza.



www.unidocs.it - Appunti e dispense per superare i tuoi esami universitari