Se σ non è noto

In genere lo scarto quadratico medio della popolazione σ, al pari della media μ, non è noto.

Pertanto, per ottenere un intervallo di confidenza per la media della popolazione, occorre utilizzare la deviazione standard del campione.

Al posto dell'errore medio $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ utilizziamo l'errore standard stimato:

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}}$$

$$s_{\bar{X}} = \frac{s}{\sqrt{n-1}} \sqrt{1 - \frac{n}{N}}$$ (per popolazioni finite)

Dove s è la deviazione standard del campione

Esercizio: stima ad intervallo con σ non noto

Su un campione di 120 intervistati si è rilevata una spesa media mensile per telefonate su cellulare di 15 euro con scarto quadratico medio di 5,4. Assumendo che la popolazione è distribuita in modo normale, stimare la spesa media nella popolazione di riferimento, con un livello di fiducia del 95,45% .

Come procedere

1. Individuare il valore di $z_{\alpha/2}$ corrispondente al livello di fiducia del 95,44%.
2. Utilizzare il valore z per costruire gli intervalli di confidenza, stimando l'errore standard mediante lo scarto quadratico medio del campione.

$$\bar{X} - Z\alpha/_{2} \cdot \frac{s}{\sqrt{n-1}} \leq \mu \leq \bar{X} + Z\alpha/_{2} \cdot \frac{s}{\sqrt{n-1}}$$

$$15 - 2\,(5.4/\sqrt{119}) \leq \mu \leq 15 + 2\,(5.4/\sqrt{119})$$

$$14{,}01 \leq \mu \leq 15{,}99$$

Possiamo dunque affermare che a partire dalla spesa media rilevata sul campione di 15 euro, la spesa media della popolazione, è compresa tra 14,01 e 15,99 euro, con un livello di fiducia del 95,45% e con una probabilità del 4,55% che sia esterna (maggiore o minore) a questo intervallo.

Esercizio

Su un campione di 110 punti vendita si è rilevato che il prezzo di vendita di un noto modello di cellulare è di 355 euro, con uno scarto quadratico medio di 16 euro.
Assumendo che la popolazione sia distribuita in modo normale, stimare il prezzo di vendita di quel prodotto nella popolazione di riferimento, con un livello di confidenza del 99,73% .

Se σ non è noto: approfondimenti

Negli esercizi precedenti in cui n era grande (n>100) , anche quando σ non era noto, abbiamo utilizzato l'errore standard stimato e abbiamo fatto riferimento, per semplicità, alla distribuzione normale standard .

In realtà, se la variabile casuale *X* ha una distribuzione normale allora la statistica :

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n-1}}}$$

ha una distribuzione *t* di Student con (n−1) **gradi di libertà**.

Una t di Student con molti gradi di libertà (n>100) si approssima ad una distribuzione normale standard.

Tuttavia per un numero inferiore di gradi di libertà e dunque al diminuire di n la distribuzione t di Student differisce da quella normale e dunque invece della variabile z si utilizza t.

T di student

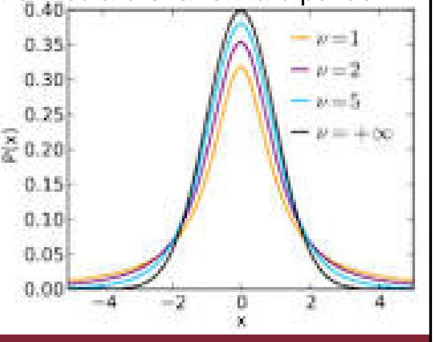
La distribuzione *t* di Student ha una forma simile a quella della normale standardizzata.

Il grafico è più appiattito e l'area sottesa sulle code è maggiore di quella della normale perché il fatto che σ non è noto e viene stimato da s, è fonte di incertezza e dunque di maggiore variabilità di *t*.

La distribuzione T è simmetrica rispetto alla media 0 e la forma dipende dal numero dei gradi di libertà

Gdl o v=(n-1)

Se n è grande la distribuzione T si approssima alla curva normale.



Intervalli di confidenza con la T di Student

gli intervalli di confidenza vengono costruiti facendo riferimento a valori di t in corrispondenza di un dato livello di confidenza e dei gradi di libertà (gdl o *v*=n-1).

Gli intervalli:

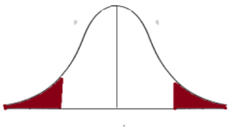
- $\bar{x} \pm t_{0,05}\frac{s}{\sqrt{n-1}}$ includono il valore incognito µ con il 95% di probabilità
- $\bar{x} \pm t_{0,01}\frac{s}{\sqrt{n-1}}$ includono il valore incognito µ con il 99% di probabilità

I valori t_{α} dipendono dal numero di gradi di libertà e vengono individuati utilizzando apposite tavole.

La tavola della T di student

Tavola distribuzione t di Student -. Bilaterale

GdL	0.50	0.40	0.20	0.10	0.05	0.025	0.01	0.005	0.001
1	1.000	1.376	3.078	6.314	**12.706**	25.452	63.657		
2	0.816	1.061	1.886	2.920	**4.303**	6.205	9.925	14.089	31.598
3	0.765	0.978	1.638	2.353	**3.182**	4.176	5.841	7.453	12.941
4	0.741	0.941	1.533	2.132	**2.776**	3.495	4.604	5.598	8.610
5	0.727	0.920	1.476	2.015	**2.571**	3.163	4.032	4.773	6.859
6	0.718	0.906	1.440	1.943	**2.447**	2.969	3.707	4.317	5.959
7	0.711	0.896	1.415	1.895	**2.365**	2.841	3.499	4.029	5.405
8	0.706	0.889	1.397	1.860	**2.306**	2.752	3.355	3.832	5.041
9	0.703	0.883	1.383	1.833	**2.262**	2.685	3.250	3.690	4.781
10	0.700	0.879	1.372	1.812	**2.228**	2.634	3.169	3.581	4.587
11	0.697	0.876	1.363	1.796	**2.201**	2.593	3.106	3.497	4.437
12	0.695	0.873	1.356	1.782	**2.179**	2.560	3.055	3.428	4.318
13	0.694	0.870	1.350	1.771	**2.160**	2.533	3.012	3.372	4.221
14	0.692	0.868	1.345	1.761	**2.145**	2.510	2.977	3.326	4.140
15	0.691	0.866	1.341	1.753	**2.131**	2.490	2.947	3.286	4.073
16	0.690	0.865	1.337	1.746	**2.120**	2.473	2.921	3.252	4.015
17	0.689	0.863	1.333	1.740	**2.110**	2.458	2.898	3.222	3.965
18	0.688	0.862	1.330	1.734	**2.101**	2.445	2.878	3.197	3.922
19	0.688	0.861	1.328	1.729	**2.093**	2.433	2.861	3.174	3.883
20	0.687	0.860	1.325	1.725	**2.086**	2.423	2.845	3.153	3.850
21	0.686	0.859	1.323	1.721	**2.080**	2.414	2.831	3.135	3.819
22	0.686	0.858	1.321	1.717	**2.074**	2.406	2.819	3.119	3.792
23	0.685	0.858	1.319	1.714	**2.069**	2.398	2.807	3.104	3.767
24	0.685	0.857	1.318	1.711	**2.064**	2.391	2.797	3.090	3.745
25	0.684	0.856	1.316	1.708	**2.060**	2.385	2.787	3.078	3.725
26	0.684	0.856	1.315	1.706	**2.056**	2.379	2.779	3.067	3.707
27	0.684	0.855	1.314	1.703	**2.052**	2.373	2.771	3.056	3.690
28	0.683	0.855	1.313	1.701	**2.048**	2.368	2.763	3.047	3.674
29	0.683	0.854	1.311	1.699	**2.045**	2.364	2.756	3.038	3.659
30	0.683	0.854	1.310	1.697	**2.042**	2.360	2.750	5.030	3.646
35	0.682	0.852	1.306	1.690	**2.030**	2.342	2.724	2.996	3.591
40	0.681	0.851	1.303	1.684	**2.021**	2.329	2.704	2.971	3.551
45	0.680	0.850	1.301	1.680	**2.014**	2.319	2.690	2.952	3.520
50	0.680	0.849	1.299	1.676	**2.008**	2.310	2.678	2.937	3.496
55	0.679	0.849	1.297	1.673	**2.004**	2.304	2.669	2.925	3.476
60	0.679	0.848	1.296	1.671	**2.000**	2.299	2.660	2.915	3.460
70	0.678	0.847	1.294	1.667	**1.994**	2.290	2.648	2.899	3.435
80	0.678	0.847	1.293	1.665	**1.989**	2.284	2.638	2.887	3.416
90	0.678	0.846	1.291	1.662	**1.986**	2.279	2.631	2.878	3.402
100	0.677	0.846	1.290	1.661	**1.982**	2.276	2.625	2.871	3.390
120	0.677	0.845	1.289	1.658	**1.980**	2.270	2.617	2.860	3.373
∞	0.675	0.842	1.282	1.645	**1.960**	2.241	2.576	2.807	3.291



La tavola fornisce i valori critici per la distribuzione t. La colonna a sinistra contiene il numero dei gradi di libertà, mentre le altre colonne danno i valori di t in corrispondenza dei vari livelli di significatività, cioè le porzioni di area nelle due code della distribuzione. Quindi α=0,050 corrisponde a due aree $\alpha/2$=0,025, a destra e a sinistra della distribuzione.

Esercizio: stima ad intervallo con σ non noto e n piccolo

Su un campione di 30 intervistati si è rilevata una spesa media mensile per sigarette elettroniche di 58 euro con scarto quadratico medio di 4 euro. Assumendo che la popolazione è distribuita in modo normale, stimare la spesa media nella popolazione di riferimento, con un livello di fiduciaa del 95% .

Come procedere
1. Calcolare α= (1-0,95)=0,050
2. Calcolare i gradi di libertà v= (n-1)
3. Cercare sulla tavola della t di Student il valore di t in corrispondenza del valore α e di v.
4. Individuare il valore di t corrispondente.
3. Utilizzare il valore t per costruire gli intervalli di confidenza

$$\overline{X} - t_\alpha \cdot \frac{s}{\sqrt{n-1}} \leq \mu \leq \overline{X} + t_\alpha \cdot \frac{s}{\sqrt{n-1}}$$

$$58 - 2{,}045\ (4/\sqrt{29}) \leq \mu \leq 58 + 2{,}045\ (4/\sqrt{29})$$

$$56{,}48 \leq \mu \leq 59{,}52$$

Possiamo dunque affermare che a partire dalla spesa media rilevata sul campione di 58 euro, la spesa media della popolazione, è compresa tra 56,48 4 59,52 euro, con un livello di fiducia del 95% e con una probabilità del 5% che sia esterna (maggiore o minore) a questo intervallo.

Esercizio

Su un campione di 25 donne si è rilevato un consumo medio di alcol settimanale di 9 unità con uno scarto quadratico medio di 2,5 unità.
Assumendo che la popolazione è distribuita in modo normale, stimare il consumo medio della popolazione di riferimento, con un livello di fiducia del 99% .

Come procedere
1. Calcolare α= (1-0,99)=0,01
2. Calcolare i gradi di libertà v= (n-1)
3. Cercare sulla tavola della t di Student il valore di t in corrispondenza del valore α e di v.
4. Individuare il valore di t corrispondente.
3. Utilizzare il valore t per costruire gli intervalli di confidenza

$$\overline{X} - t_\alpha \cdot \frac{s}{\sqrt{n-1}} \leq \mu \leq \overline{X} + t_\alpha \cdot \frac{s}{\sqrt{n-1}}$$

$$9-2{,}797\,(2{,}5/\sqrt{24}\,) \leq \mu \leq 9 + 2{,}797\,(2{,}5/\sqrt{24}\,)$$

$$7{,}57 \leq \mu \leq 10{,}42$$

Quando il parametro da stimare è una proporzione

Spesso nelle ricerche di mercato le statistiche che interessano non sono espressi in valori medi, ma in proporzioni.
Si è interessati ad esempio a conoscere la proporzione di clienti soddisfatti o insoddisfatti, oppure di consumatori di un determinato prodotto.

Una volta rilevate queste proporzioni su un campione come possiamo procedere a stimare la proporzione reale nella popolazione di riferimento?

Anche in questo caso possiamo procedere analogamente alla stima dei valori medi, poiché la distribuzione delle proporzioni campionarie p, tende, se n è grande a distribuirsi secondo una distribuzione normale, con
con media: E(p) =P
dove P è la proporzione reale nella popolazione
e varianza :

Var (p) = PQ/n dove Q=(1-P)
(popolazione, non finita con qualunque tipo di estrazione; popolazione finita con estrazione con ripetizione, n>30))

Var (p) = PQ/n ·[(N-n)/(N-1)]
(popolazione finita con estrazione senza ripetizione)

Possiamo dunque procedere analogamente a quanto abbiamo fatto per stimare i valori medi, anche nel caso di proporzioni.

Sappiamo infatti che:

- Per n grande, o per popolazioni non finite, o nell'estrazione con ripetizione:
  - il 68.26% delle proporzioni dei campioni è compreso tra $P \pm \sqrt{\frac{PQ}{n}}$
  - il 95.44% tra $P \pm 2\sqrt{\frac{PQ}{n}}$
  - il 99.73% tra $P \pm 3\sqrt{\frac{PQ}{n}}$
- Per popolazioni finite, nell'estrazione senza ripetizione:
  - il 68.26% delle proporzioni dei campioni è compreso tra $P \pm \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$
  - il 95.44% tra $P \pm 2\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$
  - il 99. 73% tra $P \pm 3\sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$

Esercizio

Su un campione di n=100 negozi, risulta che 40 hanno adottato un nuovo orario di apertura. Perciò la proporzione campionaria è di 0,40. Da altre indagini di fonte ufficiale risulta invece che la porzione di negozi in tutta la zona che hanno adottato il nuovo orario è del 36%, quindi la proporzione della popolazione è di 0,36.
Quale è la probabilità di ottenere un campione che ha una proporzione superiore di 0,40 se quella della popolazione è di 0,36?

Facendo riferimento alla distribuzione delle proporzioni campionarie la proporzione media di tutti i possibili campioni di 100 unità estraibili dalla popolazione si distribuisce normalmente con media: E(p) =P =0,36
e errore medio delle proporzioni: $\sqrt{\text{Var}(p)} = \sqrt{PQ/n} = 0{,}048$ $\quad Z=\frac{0{,}40-0{,}36}{0{,}048}= 0{,}83$

Come procedere
1. Trovare il valore medio e l'errore standard delle proporzioni campionarie
2. Calcolare il valore standardizzato
3. Disegnare la distribuzione normale
4. Calcolare la probabilità sulla tavola della distribuzione normale
5. Trarre le conclusioni

La probabilità di ottenere un campione con una proporzione
-superiore a 0,40 è di (1-0,7967)=0,2033 = 20%

Intervallo di confidenza per proporzioni

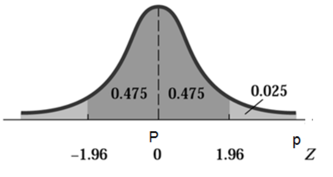
$$p - z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}} \leq P \leq p + z_{\alpha/2} \cdot \sqrt{\frac{PQ}{n}}$$

A partire dalla proporzione del campione p possiamo costruire un intervallo di valori sottraendo e sommando $z_{\alpha/2}$ e moltiplicando per l'errore .

Come sappiamo $z_{\alpha/2}$ è il valore a cui corrisponde un'area cumulata della distribuzione normale standardizzata pari a (1- α/2).

Se n è grande possiamo usare la proporzione p del campione come buona approssimazione della proporzione della popolazione nel calcolo dell'errore standard:

$$s_p = \sqrt{\frac{pq}{n}}$$



Esercizio: stima ad intervallo di una proporzione

Su un campione casuale semplice di 150 intervistati si è rilevata che la percentuale di soggetti che legge un quotidiano è del 40%.
Stimare la vera percentuale di lettori di quotidiani nella popolazione, con un livello di fiducia del 95,45% e del 99%.

Come procedere:
1. Individuare il valore di z corrispondente a livello di fiducia richiesto.
2.Utilizzare il valore z per costruire gli intervalli di confidenza

$$\boldsymbol{p - Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \leq P \leq p + Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}}$$

Attenzione p non è la percentuale, ma la proporzione!!

(1-α)=95%

$$0{,}40 - 2 \cdot \sqrt{\frac{0{,}40 \cdot 0{,}60}{150}} \leq P \leq 0{,}40 + 2 \cdot \sqrt{\frac{0{,}40 \cdot 0{,}60}{150}}$$

$$0{,}32 \leq P \leq 0{,}48$$

(1-α)=99%

$$0{,}40 - 2{,}58 \cdot \sqrt{\frac{0{,}40 \cdot 0{,}60}{150}} \leq P \leq 0{,}40 + 2{,}58 \cdot \sqrt{\frac{0{,}40 \cdot 0{,}60}{150}}$$

$$0{,}30 \leq P \leq 0{,}50$$

Possiamo dunque affermare che a partire dalla percentuale rilevata sul campione, la percentuale di lettori di quotidiani nella popolazione di riferimento è compresa tra il 32% e il 48% con un livello di fiducia del 95,45% e tra il 30% e il 50% con un livello di fiducia del 99%.

Esercizio

In un campione di 80 intervistati, 36 clienti hanno detto di preferire l'hotel Royal agli altri hotel della zona.
A- Si vuole applicare il risultato all'intera popolazione di riferimento, con un livello di confidenza del 95%. Quale intervallo di gradimento si ottiene per l'hotel Royal?
B- Se si decide di estendere la rilevazione a 250 clienti ottenendo una percentuale di preferenze per l'hotel Royal del 48%, quali sono i nuovi intervalli di confidenza?

Come procedere:
1. Calcolare la proporzione p di clienti che preferiscono l'hotel Royal
2. Calcolare l'errore standard delle proporzioni
3. Individuare il valore di z corrispondente a livello di fiducia richiesto.
4. Utilizzare il valore z per costruire gli intervalli di confidenza

$$p - Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}} \leq P \leq p + Z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$$

Risposta A

p=36/80=0,45

$$s_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0{,}45\ (1-0{,}45)}{80}} = 0{,}056$$

$$0{,}45 - 1{,}96 \cdot 0{,}056 \leq P \leq 0{,}45 + 1{,}96 \cdot 0{,}056$$

$$0{,}34 \leq P \leq 0{,}56$$

Esercizio

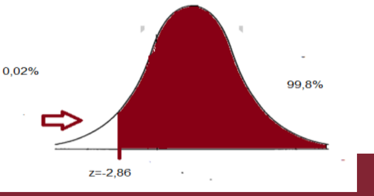
Da una ricerca di mercato effettuata su un campione di 200 intervistati risulta che solo 100 individui sono a favore della costruzione di un centro commerciale.
A- Si stimi la proporzione della popolazione a favore della costruzione calcolando l'intervallo di confidenza al 95,45%
B- Se l'impresa che costruisce il centro commerciale sostiene che nella popolazione il 60% è a favore della costruzione, qual è la probabilità di avere un campione di 200 persone con la proporzione minore o uguale a quella che abbiamo osservato se la vera proporzione della popolazione è dello 0,6 ?
L'impresa ha ragione o torto?

- Intervallo di confidenza
- 95,45% $0{,}5 \pm 2\sqrt{\frac{0{,}50\,(1-0{,}50)}{200}}$ = (0,465; 0,535)

- Probabilità di avere un campione con p≤ 0,50 se P=0,60
  - Errore standard = $\sqrt{\frac{0{,}60\,(1-0{,}60)}{200}} = 0{,}035$

$z=\frac{0{,}50-0{,}60}{0{,}035}=-2{,}86$

Pr (z≤ -2,86)= 0,02



-