

Intervalli di confidenza Inferenza I

Maura Mezzetti

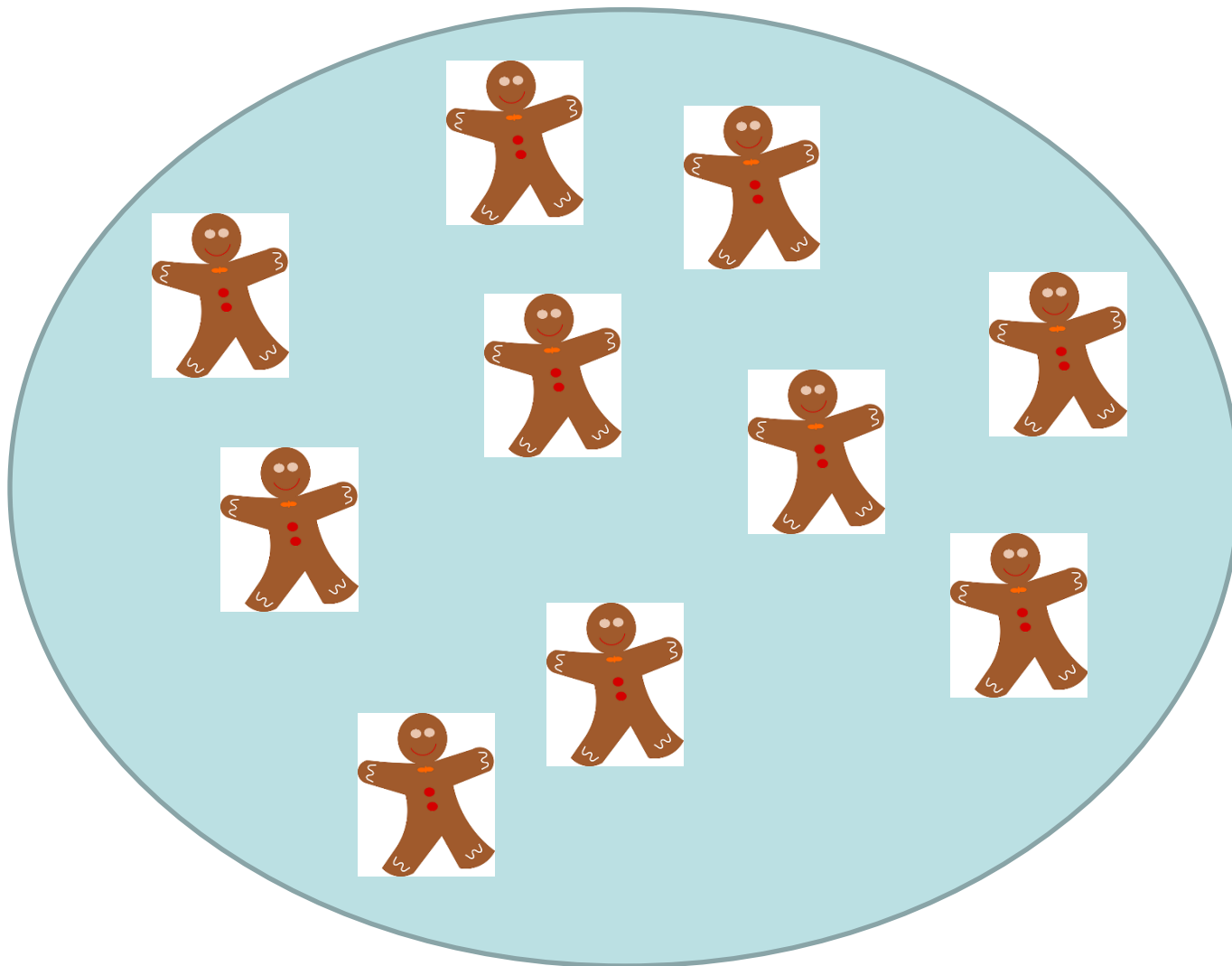
Stimatori

Si vuole stimare

- Quanti adolescenti usano regolarmente Instagram
- Quanto tempo al giorno un adolescente sta su WhatsApp

Si intervistano 10 adolescenti

Intervistiamo 10 adolescenti



Individuo	Usi Instagram?	Quanto tempo al giorno stai su WhatsApp? (in minuti)
1	Sì	10
2	No	60
3	No	100
4	No	70
5	Sì	45
6	No	80
7	No	120
8	Sì	40
9	No	55
10	Sì	45

Si vuole stimare

- Quanti adolescenti usano regolarmente Instagram
- Quanto tempo al giorno un adolescente sta su WhatsApp

Si vuole stimare

- Quanti adolescenti usano regolarmente Instagram
- Proporzione di adolescenti che usano regolarmente Instagram
- Quanto tempo al giorno un adolescente sta su WhatsApp
- Tempo medio un adolescente passa su WhatsApp

Si vuole stimare

- Proporzione di adolescenti che usano regolarmente Instagram

$Y = \text{"utilizzo/non utilizzo Instagram"}$

$Y \sim \text{Bernoulli}(p)$

Stimatore per p ?

Stima per p ?

$Y \sim \text{Bernoulli}(p)$

Stimatore per p ?

Stima per p ?

$Y \sim \text{Bernoulli}(p)$

Stimatore per p ?

Proporzione campionaria

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$$

Stima per p ?

$$\hat{p} = \frac{4}{10} = 0.4$$

Si vuole stimare

- Tempo medio un adolescente passa su WhatsApp

$Y = \text{"tempo su WhatsApp"}$

$$E(Y) = \mu$$

Stimatore per μ ?

Stima per μ ?

$$E(Y)=\mu$$

Stimatore per μ ?

Stima per μ ?

$$E(Y)=\mu$$

Stimatore per μ ?

Media Campionaria

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Stima per μ ?

$$\bar{x} = \frac{\sum_{i=1}^n X_i}{n}$$

$$\bar{x} = \frac{10 + 60 + 100 + 70 + 45 + 80 + 120 + 40 + 55 + 45}{10} = 62.5$$

Stimatori

- Quanti adolescenti usano regolarmente Instagram **40%**
- Quanto tempo al giorno un adolescente sta su WhatsApp **62.5 minuti al giorno**

Quanto precise sono le precedenti stime?

Si vuole stimare

- Proporzione di adolescenti che usano regolarmente facebook

Y ="utilizzo/non utilizzo facebook"

$Y \sim \text{Bernoulli}(p)$

Stimatore per p ?

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$$

Varianza dello stimatore

$$\text{Var}(\hat{p}) = \frac{p(1-p)}{n}$$

- Proporzione di adolescenti che usano regolarmente facebook

Y="utilizzo/non utilizzo facebook"

Y Bernoulli(p)

Stimatore per p ?

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$$

Varianza dello stimatore

$$Var(\hat{p}) = \frac{p(1-p)}{n}$$

Stima della varianza dello stimatore

$$Var(\hat{p}) = \frac{0.4 \times 0.6}{10} = 0.024$$

Tempo medio un adolescente passa su WhatsApp

$Y =$ "tempo medio su WhatsApp"

$$E(Y) = \mu$$

Stimatore per μ

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Stima per μ $\bar{x} = 62.5$

Varianza dello stimatore

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$\frac{\sum_{i=1}^n X_i^2}{n-1} = \frac{10^2 + 60^2 + 100^2 + 70^2 + 45^2 + 80^2 + 120^2 + 40^2 + 55^2 + 45^2}{9}$$

$$\frac{\sum_{i=1}^n X_i^2}{n-1} = \frac{48075}{9} = 5341.73$$

$$S^2 = \frac{\sum_{i=1}^n X_i^2}{n-1} - \frac{n}{n-1} \bar{X}^2 = 5341.73 - \frac{10}{9} 3906.35 = 1001.3$$

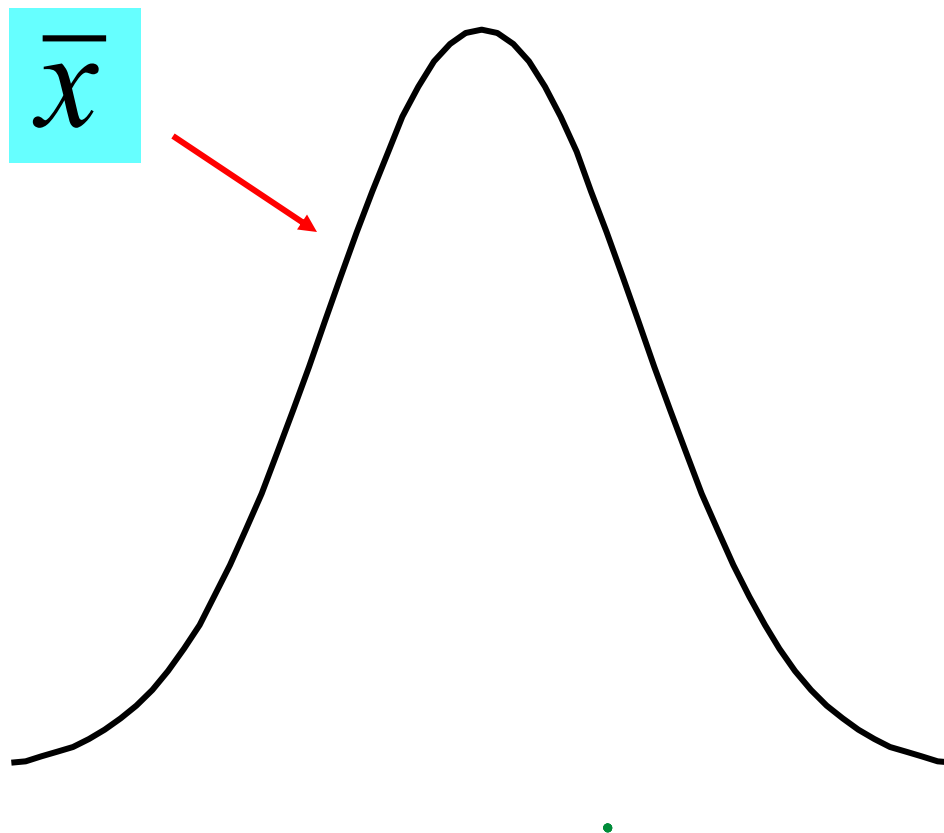
$$\widehat{Var}(\bar{X}) = \frac{S^2}{10} = \frac{1001.3}{10} = 100.13$$

Introduzione

La stima puntuale va corredata da una misura della sua precisione. La varianza dello stimatore serve proprio a questo. Quanto più questa è piccola tanto più lo stimatore può dirsi preciso. Un modo rapido ed efficace di comunicare l'incertezza ed il grado di precisione connessi ad una stima è attraverso un intervallo di valori che dia garanzia di contenere il valore incognito del parametro con una determinata affidabilità.

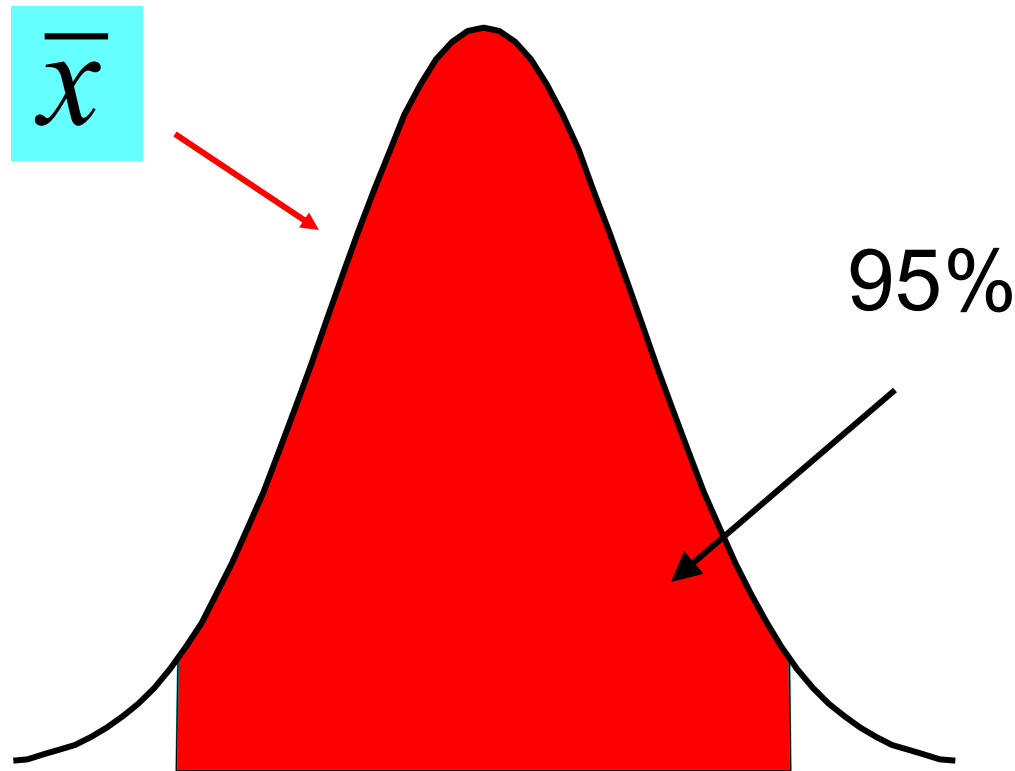
Sia $X \sim N(\mu, \sigma^2)$ σ^2 parametro noto e μ parametro incognito. Stimiamo μ tramite la media campionaria.

$\bar{X}$



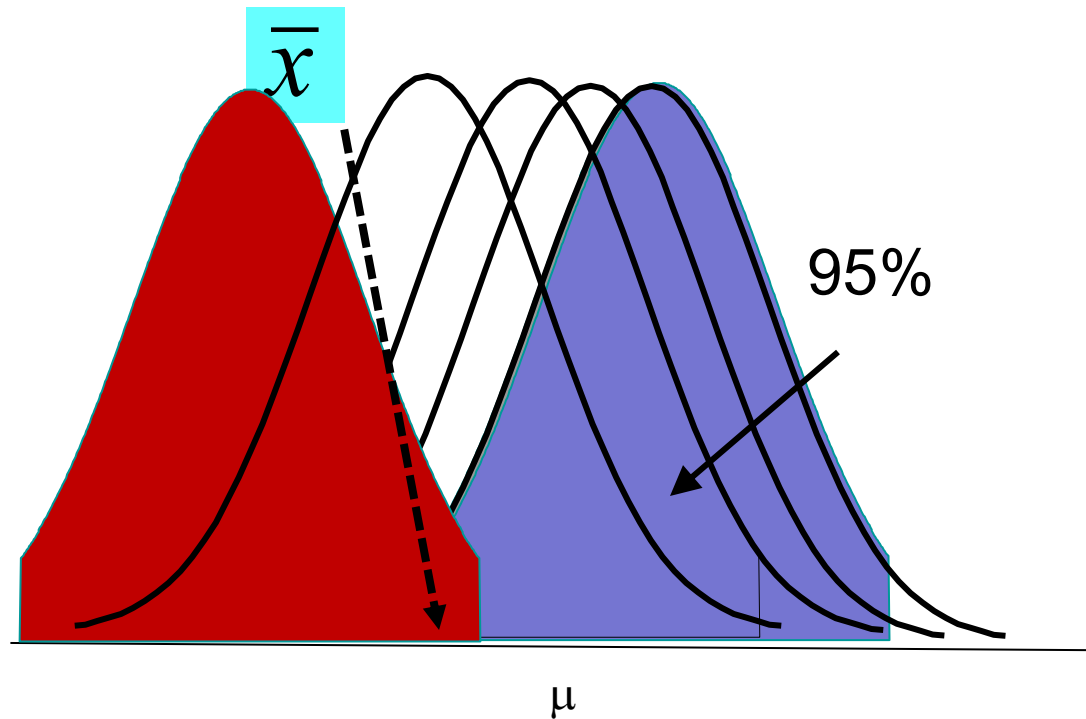
Sia $X \sim N(\mu, \sigma^2)$ σ^2 parametro noto e μ parametro incognito. Stimiamo μ tramite la media campionaria.

$\bar{X}$



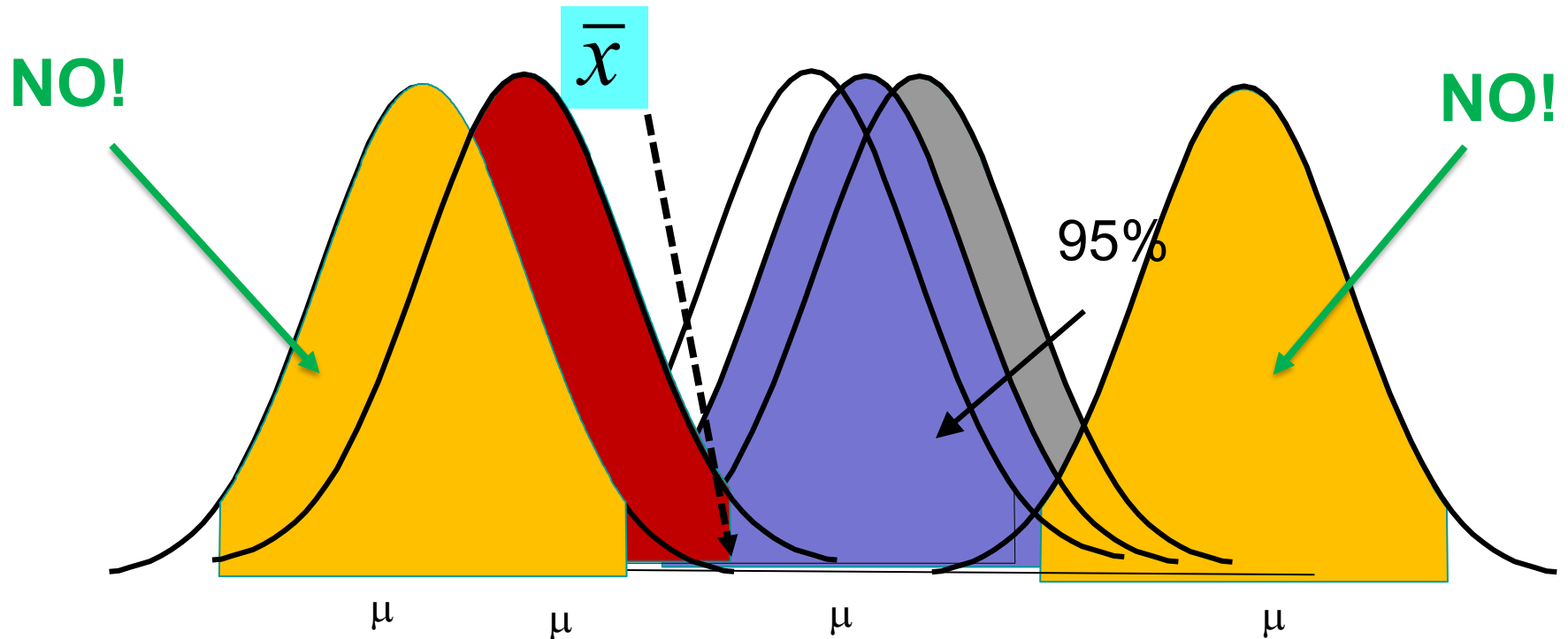
$X \sim N(\mu, \sigma^2)$ σ^2 known parameter and μ unknown parameter.

We observe $\bar{x}$ can we guess what was the original distribution?



$X \sim N(\mu, \sigma^2)$ σ^2 known parameter and μ unknown parameter.

We observe $\bar{x}$ can we guess what was the original distribution?



$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{Normale standard}$$

$$P(-1.96 \leq Z \leq 1.96) = 0.95$$

$$P\left(-1.96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) = P\left(\mu - 1.96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$\text{Probabilità } \bar{X} \in \text{all'intervallo } \left(\mu - 1.96 \frac{\sigma}{\sqrt{n}}; \mu + 1.96 \frac{\sigma}{\sqrt{n}}\right) \text{ è } 0.95$$

Probabilità che l'intervallo aleatorio $\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right)$ contenga il parametro μ è 0.95

95% Intervallo di Confidenza aleatorio per μ è

$$\left(\bar{X}_n - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{X}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

Il 95% intervallo di confidenza numerico si ottiene sostituendo i valori osservati alle variabili aleatorie

$$\left(\bar{x}_n - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{x}_n + 1.96 \frac{\sigma}{\sqrt{n}} \right)$$

L'intervallo di confidenza a livello $(1-\alpha)$ per la media della popolazione, nel caso di popolazione normale con varianza nota

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

$1-\alpha$	α	$z_{1-\alpha/2}$
0.90	0.1	1.645
0.95	0.05	1.96
0.99	0.01	2.576

Esempio: stima della media di una pop. normale, varianza nota

Si assuma che il peso di un prodotto possa essere descritto da una v.c. normale con media incognita e deviazione standard (σ) pari a 3 grammi, $X \sim N(\mu, 9)$.

Viene selezionato un campione casuale di 20 unità, $\{X_1, \dots, X_{20}\}$.
Lo stimatore media campionaria ha la seguente distribuzione:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \sim N(\mu, 0.45)$$

dove $\frac{\sigma^2}{n} = \frac{9}{20} = 0.45$. Equivalentemente,

$$\frac{\bar{X} - \mu}{\sqrt{0.45}} \sim N(0, 1)$$

Ciò implica che

$$P \left(-1.96 \leq \frac{\bar{X} - \mu}{\sqrt{0.45}} \leq 1.96 \right) = 0.95$$

ovvero che

$$P \left(\mu - 1.96\sqrt{0.45} \leq \bar{X} \leq \mu + 1.96\sqrt{0.45} \right) =$$

$$P \left(\mu - 1.31 \leq \bar{X} \leq \mu + 1.31 \right) = 0.95$$

Scambiando il ruolo di $\bar{X}$ e μ nell'espressione precedente,

$$P \left(\bar{X} - 1.31 \leq \mu \leq \bar{X} + 1.31 \right) = 0.95$$

Si osservi infatti che

$$P \left(\left| \frac{\bar{X} - \mu}{\sqrt{0.45}} \right| \leq 1.96 \right) = P \left(\left| \frac{\mu - \bar{X}}{\sqrt{0.45}} \right| \leq 1.96 \right) = 0.95$$

L'intervallo di valori $[\bar{X} - 1.31, \bar{X} + 1.31]$ rappresenta l'intervallo di confidenza al 95% per la media μ .

La media campionaria calcolata sul campione osservato risulta pari a $\bar{x}=99$ grammi (che rappresenta la nostra stima puntuale della media).

L'intervallo di confidenza stimato risulta dunque:

$$[99 - 1.31, 99 + 1.31] = [97.69, 100.31]$$

Definizione di intervallo di confidenza

Sia $X \sim f(x; \theta)$, e sia $(X_1, X_2, \dots, X_n)$ un campione iid, due statistiche T_1 e T_2 tali che

1. $T_1 < T_2$
2. $P(T_1 < \theta < T_2) = 1 - \alpha \quad \forall \theta \in \Theta$

(T_1, T_2) è intervallo di confidenza a livello $(1 - \alpha)$

(t_1, t_2) è l'intervallo di confidenza numerico a livello $(1 - \alpha)$

Proprietà dell'intervallo di confidenza (CI)

- La statistica campionaria (in questo caso la media) è il punto centrale dell'intervallo.
- Maggiore è il livello di confidenza, più largo è l'intervallo associato
- All'aumentare dell'ampiezza campionaria diminuisce l'ampiezza dell'intervallo

L'intervallo di confidenza viene scritto:

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$$

**Media
campionaria**

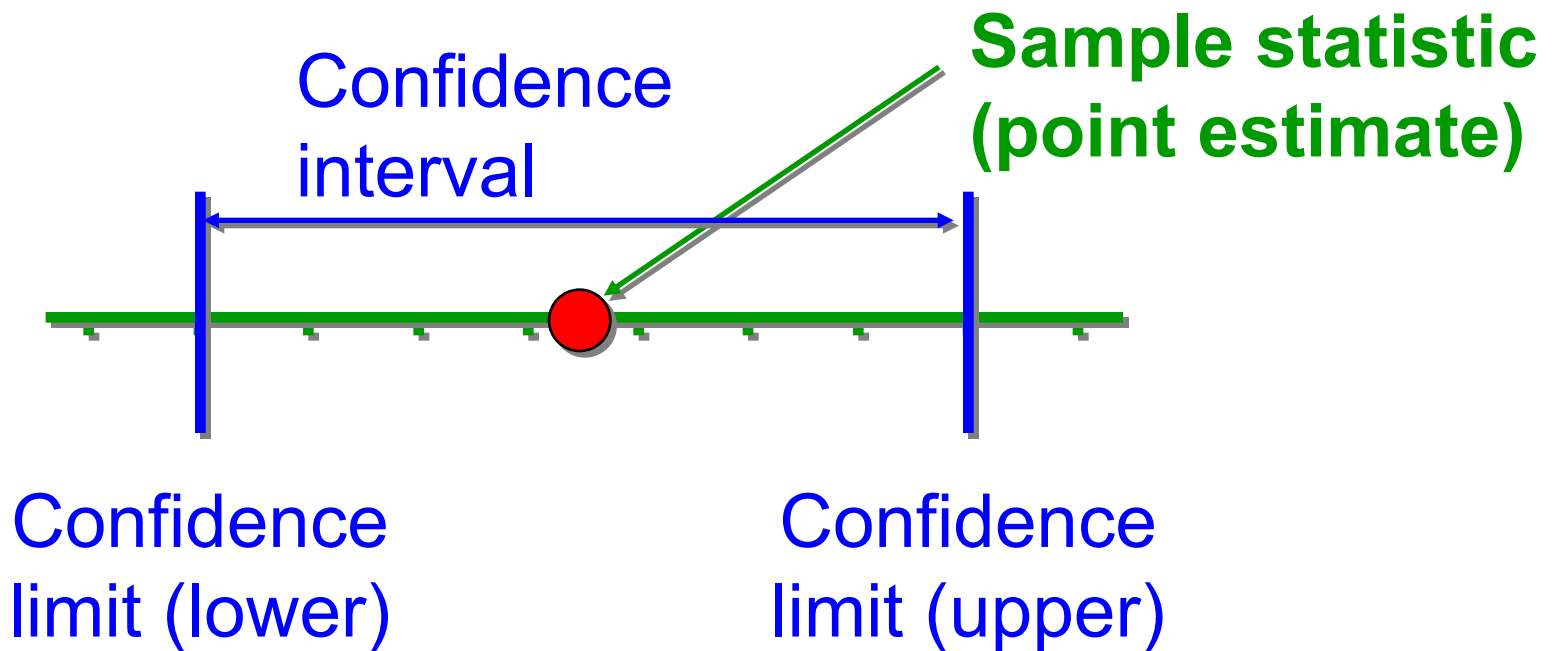
```
graph LR; A[Media campionaria] --- B["x̄"]; C[Livello di confidenza] --- D["1.96"]; E["(standard error della media)"] --- F["σ / √n"]
```

**Livello di
confidenza**

**(standard error
della media)**

Livello di confidenza

Probabilità che un dato intervallo contenga il parametro incognito della popolazione

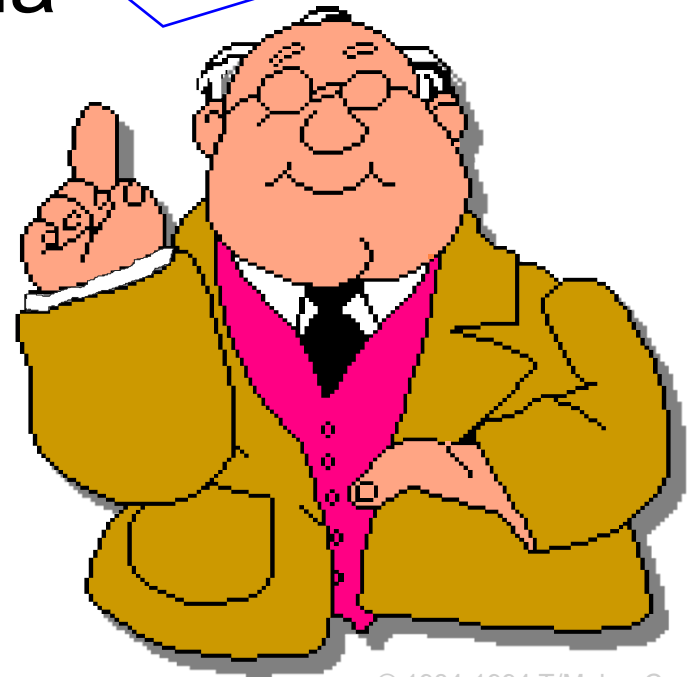


Da cosa dipende lunghezza intervallo di confidenza?

- 1. Dispersione dei dati
 - Misurata da σ
- 2. Ampiezza campionaria
 - $\sigma_{\bar{X}} = \sigma / \sqrt{n}$
- 3. Livello di confidenza
 - $(1 - \alpha)$

Intervallo di confidenza

$$\bar{X} - Z\sigma_{\bar{X}} ; \bar{X} + Z\sigma_{\bar{X}}$$



Esercizio

Si consideri il seguente campione di dimensione $n = 8$ estratto nell'ambito di un'indagine sulle condizioni lavorative dei laureati in economia dopo il 2000:

Sulla base di tale campione: si stimi l'età media dei laureati dopo il 2000 e si costruisca un intervallo di confidenza al 99% per μ sotto l'ipotesi che l'età ha distribuzione normale con varianza

$$\sigma^2 = 30$$

Unità	1	2	3	4	5	6	7	8
Sesso	F	F	M	F	M	F	M	F
Età	33	25	41	37	27	30	26	35
Lavoratore	S	S	S	N	S	N	S	N

$$\bar{x} = \frac{1}{8} \sum_{i=1}^n x_i = \frac{1}{8} (33 + 25 + \dots + 35) = \frac{254}{8} = 31,57$$

$$P\left(\bar{X} - z_{\alpha/2} \sqrt{\sigma^2 / n} < \mu < \bar{X} + z_{\alpha/2} \sqrt{\sigma^2 / n}\right) = 1 - \alpha$$

$$(31,57 - 2,57 \sqrt{30 / 8}; 31,57 + 2,57 \sqrt{30 / 8};) \\ = (26,78; 36,73)$$

Esercizio

Il Comune di Roma deve valutare se sospendere la circolazione delle auto la prossima domenica. Ai fini della decisione, dispone delle rilevazioni del livello delle polveri fini nell'aria effettuate in 240 centraline. Se il livello delle polveri è superiore a $50 \text{ } (\mu\text{g}/\text{m}^3)$, si rende opportuna la limitazione della circolazione delle auto.

Il livello delle polveri fini nell'aria viene descritto da una variabile aleatoria X , con valore atteso $E(X) = \mu$ (“livello medio delle polveri nell'aria”) e varianza uguale a 1080. Avendo osservato sulle 240 centraline un valore medio uguale a 56:

- proporre uno stimatore per μ (“livello medio delle polveri nell'aria”);

Stimatore migliore per μ è la media campionaria

$$\bar{X} = \frac{1}{n} \sum x_i$$

- descrivere le proprietà dello stimatore utilizzato in a);

$$\bar{X} = \frac{1}{n} \sum x_i$$

è stimatore corretto, efficiente e consistente per μ

- costruire un intervallo di confidenza per μ al livello del 95%;

**IC per μ con distribuzione non normale e varianza nota.
Sfruttando TLC,**

$$95\%CI = \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad z_{1-\alpha/2} = z_{1-0.05/2} = z_{0.975} = 1.96$$

$$95\%CI = \left[56 - 1.96 \sqrt{\frac{1080}{240}}, 56 + 1.96 \sqrt{\frac{1080}{240}} \right] =$$

$$[51.842; 60.1577]$$

- descrivere come si modifica l'intervallo in
c) se, a parità di condizioni, aumentiamo il
livello di confidenza



- descrivere come si modifica l'intervallo in
c) se, a parità di condizioni, aumentiamo il
livello di confidenza

L'ampiezza dell'IC cresce.

**Infatti, aumentando $1-\alpha$, α si riduce, facendo
aumentare l'ampiezza dell'IC**

$X \sim N(\mu, \sigma^2)$ μ, σ^2 non nota

RICORDIAMOCI come stimare la varianza della popolazione σ^2 :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

S^2 è uno stimatore distorto!!!

S^2_c è uno stimatore corretto!!!

$$S_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$S_c^2 = \frac{1}{n-1} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right)$$

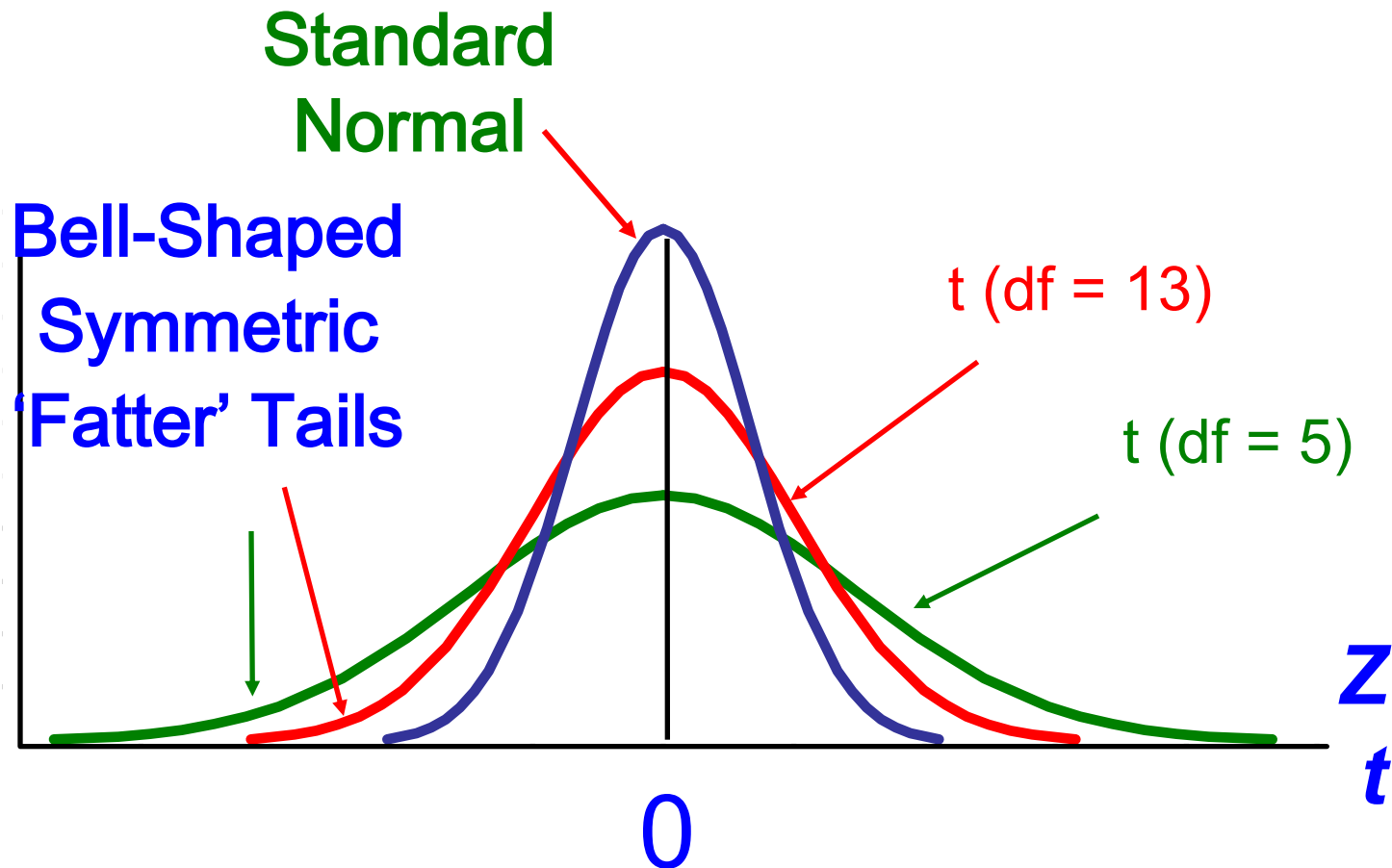
La costruzione precedente dell'intervallo di confidenza si basava sulla distribuzione Normale standard:

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$T_n = \frac{\bar{X}_n - \mu}{\frac{S_C}{\sqrt{n}}} \quad \text{non è una Normale standard!!!}$$

T_n si distribuisce come una distribuzione t-student's con $(n-1)$ -gradi di libertà

Student's t Distribution



Distribuzione Student's t

La Distribuzione T-Student ha una forma campanulare, simmetrica e dipende da un parametro indicato con n .

All'aumentare di n , la distribuzione è sempre più vicina alla distribuzione Gaussiana

k	Valori della funzione di ripartizione della distribuzione $T^{(k)}$						
	0.75	0.9	0.95	0.975	0.99	0.995	0.9995
1	1.000	3.078	6.314	12.706	31.821	63.656	636.578
2	0.816	1.886	2.920	4.303	6.965	9.925	31.600
3	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.689
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.660
30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	1.299	1.676	2.009	2.403	2.678	3.496
60	0.679	1.296	1.671	2.000	2.390	2.660	3.460
70	0.678	1.294	1.667	1.994	2.381	2.648	3.435
80	0.678	1.292	1.664	1.990	2.374	2.639	3.416
90	0.677	1.291	1.662	1.987	2.368	2.632	3.402
100	0.677	1.290	1.660	1.984	2.364	2.626	3.390
120	0.677	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.675	1.282	1.645	1.960	2.327	2.576	3.291

Intervallo di confidenza (σ non noto)

$$\bar{X} - t_{\alpha/2, n-1} \times \frac{S_C}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1} \times \frac{S_C}{\sqrt{n}}$$

Percentili della v.c. t di Student per $n = 20$ (confronto con la normale standard)

α	$z_{\alpha/2}$	$t_{\alpha/2}$
0.10	1.64	1.72
0.05	1.96	2.09
0.01	2.58	2.84

Esercizio

Una fabbrica produce penne stilografiche.

Una prova su 144 penne stilografiche scelte casualmente ha indicato un peso medio di 80 grammi.

La distribuzione del peso di una penna stilografica è arbitraria, di parametri μ non nota e $\sigma^2 = 16$. Si determini l'intervallo di confidenza per μ al livello $(1-\alpha) = 0.90$.

Soluzione

Intervallo di confidenza per la media al livello 90 %

$$\bar{x} \pm z_{(1-\alpha)/2} \times \sigma / \sqrt{n}$$

$$80 \pm 1.645 \times 4 / \sqrt{144}$$

$$80 \pm 0.5483$$

$$79.4517; 80.5483$$

Esempio

Un'impresa desidera stimare il tempo medio in giorni lavorativi che occorre per evadere un ordine. Per questo ha estratto un campione casuale di 12 ordini. Supponiamo X sia distribuita come una Gaussiana

$$\{x_1, \dots, x_{12}\} = \{20, 10, 15, 8, 10, 12, 9, 12, 14, 8, 13, 16\}$$

da cui risulta $\bar{x} = 12.25$ e $s^2 = 12.93$.

L'i.c. stimato al 95% per la media della popolazione è

$$12.25 \pm 2.20 \sqrt{12.93/12} = [9.97, 14.53]$$

Esercizio

La fabbrica A produce matite colorate.

Una prova su 100 matite scelte a caso ha indicato un peso medio di 12 grammi. Supponendo che il peso di una matita sia una variabile aleatoria distribuita secondo la legge normale, di parametri μ non nota e $\sigma^2 = 4$.

- a) Si proponga uno stimatore consistente per μ e si dica quanto vale la sua varianza.
- b) Si determini l'intervallo di confidenza per μ al livello $(1-\alpha) = 0.95$.

a) STIMATORE: media campionaria

b) VARIANZA: $\sigma^2/n=4/100=0.04$

$$\begin{aligned} 95\%CI &= \left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}} \right] = \\ &= [12 - 1.96 \times 0.2; 12 + 1.96 \times 0.2] = [11.608, 12.392] \end{aligned}$$

Determinazione della numerosità campionaria

Consideriamo il caso della stima della media di una popolazione normale con varianza nota.

Vogliamo determinare la numerosità campionaria n che garantisce, con un livello di confidenza pari a $1 - \alpha$, che l'errore di stima non superi un certo valore assegnato, δ :

$$P(|\bar{X} - \mu| \leq \delta) = 1 - \alpha$$

$$P(\bar{X} - \delta \leq \mu \leq \bar{X} + \delta) = 1 - \alpha$$

Dovendo essere $\delta = z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$, si ricava

$$n = z_{\alpha/2}^2 \frac{\sigma^2}{\delta^2}$$

Esercizio

Una fabbrica produce barrette di cioccolato. Il peso di ciascuna barretta, X è un numero aleatorio distribuito secondo la legge normale di parametri μ non nota e $\sigma^2 = 64$. Si consideri un campione bernoulliano di 25 barrette, per cui si ottiene $\bar{x} = 98$

- a) Si determini l'intervallo di confidenza per μ al livello $(1-\alpha) = 0.90$.
- b) Si determini l'ampiezza campionaria minima n^* che assicura che la lunghezza dell'intervallo di confidenza per μ al livello $(1-\alpha) = 0.90$ sia inferiore a 10.

$$\bar{x} = 98$$

$$90\%CI = \left[\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}} \right] = \left[98 - 1.645 \times \frac{8}{5}; \bar{x} + 1.645 \times \frac{8}{5} \right] =$$

$$= [95.368, 100.632]$$

$$\bar{x} = 98$$

$$90\%CI = \left[\bar{x} - 1.645 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1.645 \frac{\sigma}{\sqrt{n}} \right] = \left[98 - 1.645 \times \frac{8}{5}; \bar{x} + 1.645 \times \frac{8}{5} \right] =$$

$$= [95.368, 100.632]$$

$$2 \times 1.645 \times \frac{8}{\sqrt{n}} \leq 10$$

$$\sqrt{n} \geq 2 \times 1.645 \times \frac{8}{10} = 2.632$$

N almenno 7

Intervallo di confidenza per una proporzione

$X \sim \text{Bernoulli}(\pi)$

Campione casuale di dimensione n :

$\{X_1, \dots, X_n\}$, $X_i \sim \text{IID Bernoulli}(\pi)$

Campione osservato: $\{x_1, \dots, x_n\}$

Consideriamo la frequenza relativa dei successi nel campione: $\bar{X}$.

Sappiamo che $E(\bar{X}) = \pi$ e

$$V(\bar{X}) = \pi(1 - \pi)/n.$$

Se n è sufficientemente grande possiamo invocare il teorema del limite centrale ed assumere che $\bar{X} \sim N(\pi, V(\bar{X}))$. Tuttavia, la varianza della distribuzione dipende da π .

Esistono due soluzioni per questo problema. La prima consiste nell'assumere che

$$\bar{X} \sim N(\pi, \hat{V}(\bar{X})), \quad \hat{V}(\bar{X}) = \bar{X}(1 - \bar{X})/n$$

dove $\hat{V}(\bar{X})$ costituisce lo stimatore della varianza di $\bar{X}$.
L'intervallo di confidenza per π è:

$$\left[\bar{X} - z_{\alpha/2} \sqrt{\hat{V}(\bar{X})}, \bar{X} + z_{\alpha/2} \sqrt{\hat{V}(\bar{X})} \right]$$

mentre l'intervallo stimato è

$$\left[\bar{x} - z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}}, \bar{x} + z_{\alpha/2} \sqrt{\frac{\bar{x}(1 - \bar{x})}{n}} \right]$$

Esempio

Mediante un exit poll si desidera stimare la percentuale di voti di un candidato. Viene selezionato un campione casuale di $n = 1200$ unità; 622 hanno dichiarato di aver votato per quel candidato. Pertanto, $100 \times \bar{x} = 51.83\%$. Si può avere la ragionevole certezza di essere eletti?

Ponendo $\alpha = 0.01$, l'intervallo di confidenza per π al 99% ha estremi

$$\left[0.5183 - 2.58 \sqrt{\frac{0.5183(1-0.5183)}{1200}}, 0.5183 + 2.58 \sqrt{\frac{0.5183(1-0.5183)}{1200}} \right] = [0.4811, 0.5555]$$

Intervallo di confidenza per la media della popolazione μ

- $X \sim N(\mu, \sigma^2)$ σ^2 valore noto

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- $X \sim N(\mu, \sigma^2)$ σ^2 valore NON noto

$$\left(\bar{x}_n - t_{1-\alpha/2, n-1} \frac{s_c}{\sqrt{n}} ; \bar{x}_n + t_{1-\alpha/2, n-1} \frac{s_c}{\sqrt{n}} \right) \quad s_c^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Intervallo di confidenza per la media della popolazione μ

- X distribuzione arbitraria $E(X)=\mu$

$\text{Var}(X)=\sigma^2$ valore noto – n grande

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right)$$

- σ^2 valore NON noto - n grande

$$\left(\bar{x}_n - z_{1-\alpha/2} \frac{s_c}{\sqrt{n}} ; \bar{x}_n + z_{1-\alpha/2} \frac{s_c}{\sqrt{n}} \right) \quad s_c^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Intervallo di confidenza per p proporzione della popolazione

- $X \sim \text{Bernoulli}(p)$
- $E(X)=p$ $\text{Var}(X)=p(1-p)$ n grande

Intervallo di confidenza per p proporzione della popolazione

- $X \sim \text{Bernoulli}(p)$
- $E(X)=p$ $\text{Var}(X)=p(1-p)$ n grande

$$\left(\hat{p} - z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z_{1-\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

Esercizio

Una ricerca di mercato è interessata a sapere quando, e quanto spesso gli abitanti di una piccola città guardano la televisione. Un campione di 120 abitanti è selezionato e ad ognuno è stato chiesto di indicare esattamente quando hanno guardato la televisione in una data settimana. E' risultato che:

50 guardano il telegiornale serale almeno due sere la settimana

Sia X_i il tempo (in ore) passato davanti alla televisione alla settimana dell' i -esimo intervistato

$$\sum_{i=1}^{120} x_i = 1680 \quad \sum_{i=1}^{120} x_i^2 = 24600$$

Proporre uno stimatore non distorto per μ ,
la media di X e valutare la stima corrispondente

STIMATORE μ

STIMA PER μ

Proporre uno stimato non distorto per σ^2 la varianza di X
e valutare la stima corrispondente

STIMATORE σ^2

STIMA PER σ^2

Proporre uno stimatore non distorto per μ ,
la media di X e valutare la stima corrispondente

Estimator for μ

Sample mean

Estimate for μ

$$\bar{x} = \frac{\sum_{i=1}^{120} 1680}{120} = 14$$

Proporre uno stimato non distorto per σ^2 la varianza di X
e valutare la stima corrispondente

Estimator for σ^2

$$S_C^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

Estimate for σ^2

$$s_c^2 = \frac{1}{119} \left(\sum_{i=1}^{120} x_i^2 - 120 \bar{x}^2 \right) = \frac{24600 - 120 \times 14^2}{119} = 9.076$$

Quantificare la varianza dello stimatore proposto al punto 1)

Proporre un intervallo di confidenza al 99% per μ

Confidence Interval (CI) equation			
99% numerical CI			

Quantificare la varianza dello stimatore proposto al punto 1)

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$\hat{Var}(\bar{X}) = \frac{s_c^2}{n} = \frac{9.076}{120} = 0.076$$

Proporre un intervallo di confidenza al 99% per μ

Confidence Interval (CI)
equation

$$99\%CI = \left[\bar{x} - t_{0.05}^{n-1} \frac{s_c}{\sqrt{n}}, \bar{x} + t_{0.05}^{n-1} \frac{s_c}{\sqrt{n}} \right]$$

$$99\%CI = \left[\bar{x} - z_{0.995} \frac{s_c}{\sqrt{n}}, \bar{x} + z_{0.995} \frac{s_c}{\sqrt{n}} \right]$$

99% numerical CI

$$99\%CI = \left[14 - 2.576 \sqrt{\frac{9.076}{120}}, 14 + 2.576 \sqrt{\frac{9.076}{120}} \right]$$

$$99\%CI = [13.2916, 14.7084]$$

Proporre un intervallo di confidenza al 95% per p ,
la proporzione di abitanti che guardano il tg serale
almeno due sere la settimana

Confidence Interval (CI) equation for p	
95% numerical CI for p	

Proporre un intervallo di confidenza al 95% per p,
la proporzione di abitanti che guardano il tg serale
almeno due sere la settimana

Confidence Interval (CI) equation
for p

$$95\%CI = \left[\hat{p} - z_{0.975} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}, \hat{p} + z_{0.975} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \right]$$

95% numerical CI for p

$$95\%CI = \left[0.4167 - 1.96 \sqrt{\frac{0.4167 \times 0.5833}{120}}, 0.4167 + 1.96 \sqrt{\frac{0.4167 \times 0.5833}{120}} \right]$$

$$95\%CI = [0.3285, 0.5049]$$

Esercizio

Sia X il tempo impiegato da un macchinario per compiere un intero ciclo di produzione. Il fornitore del macchinario sostiene che X si distribuisce come una Normale con media 20 minuti e varianza 10 minuti².

Il potenziale acquirente del macchinario però sospetta che il tempo medio di completamento del ciclo produttivo sia maggiore di 20.

Costruire l'intervallo di confidenza per il tempo di produzione ad un livello di confidenza del 99% supponendo di trovare una media di 22 su un campione di ampiezza 9

Soluzione

$$99\%CI = \left[\bar{x} - z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} ; \bar{x} + z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}} \right] =$$
$$\left[22 - 2.58 \sqrt{\frac{10}{9}} ; 22 + 2.58 \sqrt{\frac{10}{9}} \right] = [19.291; 24.709]$$

Esercizio

In una popolazione il peso corporeo è distribuito come una normale con media μ e varianza σ^2 non note. Estratto un campione di 101 individui, si è osservato un valore medio del peso corporeo uguale a 80kg ed una stima della varianza $S^2 = 36$.

Soluzione

- proporre uno stimatore per μ (“valore atteso peso”);

media campionaria

$$\bar{X} = \frac{1}{n} \sum x_i$$

- descrivere le proprietà dello stimatore utilizzato in a);

$\bar{X}$ è stimatore corretto, efficiente e consistente per μ

Soluzione

- costruire un intervallo di confidenza per μ al livello del 90%;

IC per μ con distribuzione normale e varianza non nota →

$$90\%CI = \left[\bar{x} \pm t_{1-\alpha/2}^{n-1} \frac{s}{\sqrt{n}} \right]$$

Dalle tavole

$$t_{\alpha/2}^{n-1} = t_{0.1/2}^{100} = 1.6602$$

Quindi =

$$90\%CI = \left[80 \pm 1.6602 \sqrt{\frac{36}{101}} \right] = [80 \pm 0.99118]$$

$$[79.009; 80.991]$$