STATISTICA INFERENZIALE – SCHEDA N. 2
INTERVALLI DI CONFIDENZA PER IL VALORE ATTESO E LA FREQUENZA

Nella scheda precedente abbiamo visto come si stima un parametro incognito di una variabile aleatoria definita su una popolazione; in particolare abbiamo considerato la cosiddetta **stima puntuale** della media e della frequenza relativa, dove "stima puntuale" significa fornire **un** valore per il parametro, usando dati campionari.
In questa scheda costruiremo un intervallo nel quale ci aspettiamo stia il parametro da stimare con un elevato grado di fiducia. Questa "fiducia" è assegnata in termini probabilistici e viene detta confidenza (con una cattiva traduzione dall'inglese confidence).
Un tale intervallo si dice **intervallo di confidenza** e la probabilità (che indicheremo con 1-α) assegnata viene detta **livello di significatività** (o livello di confidenza). Usualmente si sceglie come livello di significatività il 95% o il 99%.

1. Intervalli di confidenza per il valore atteso

Se si vuole stimare la media μ di una variabile aleatoria X definita su una popolazione tramite un campione di numerosità fissata, allora si può scegliere come stimatore $\overline{X}_n$.
Un esempio è la stima del prezzo medio di un litro di latte in Liguria. Qui la popolazione è formata dai prezzi di tutti i litri di latte venduti in un determinato periodo in Liguria. Per determinare il prezzo medio l'ISTAT (Istituto Nazionale di Statistica) effettua un campionamento su vari negozi della regione, tenendo conto della dislocazione geografica, del tipo di distribuzione (supermercato, negozio) e di altri fattori. Nella nostra indagine sui prezzi di alcuni prodotti delle nostre zone di residenza abbiamo effettuato un campionamento non molto rappresentativo: comunque utilizzeremo questi dati e poi li confronteremo con quelli ufficiali.

Una stima puntuale del valore atteso μ è data dal valore $\overline{x}_n$ assunto dalla variabile $\overline{X}_n$ nel campione.
Un *intervallo di confidenza*, a livello di significatività del 95%, è un intervallo aleatorio

$$\left(\overline{X}_n - \delta, \overline{X}_n + \delta\right)$$

con δ scelto in modo tale che

$$P\left(\overline{X}_n - \delta < \mu < \overline{X}_n + \delta\right) = 0.95,$$

ossia tale che la probabilità di sbagliare sia pari a α=0.05 e quindi bassa.
La realizzazione campionaria dell'intervallo è:

$$\left(\overline{x}_n - \delta, \overline{x}_n + \delta\right)$$

Come si calcola δ?
Il calcolo dell'intervallo di confidenza si basa sulla probabilità che la variabile aleatoria $\overline{X}_n$ sia compresa fra $\mu-\delta$ e $\mu+\delta$: $\quad 0.95 = P\left(\mu - \delta < \overline{X}_n < \mu + \delta\right)$

È quindi necessario *conoscere la distribuzione di probabilità dello stimatore*. Questo è possibile se si conosce la distribuzione della variabile aleatoria X di partenza.

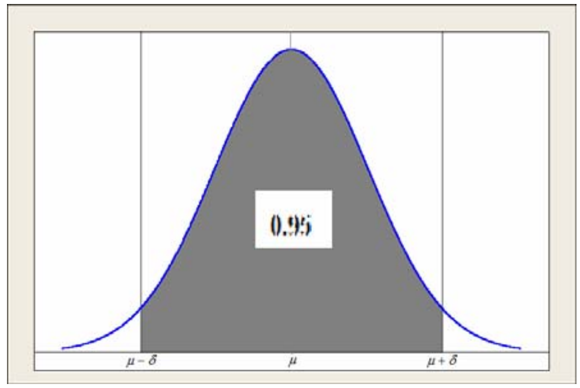
In particolare, se X ha distribuzione normale, anche $\overline{X}_n$ ha distribuzione normale con valore atteso μ e sappiamo calcolare δ in modo che:

$$P\left(\mu-\delta<\overline{X}_n<\mu+\delta\right)=0.95$$

Questo valore di δ ci permette di trovare l'intervallo di confidenza. Infatti:

$$P\left(\mu-\delta<\overline{X}_n<\mu+\delta\right) = P\left(\overline{X}_n-\delta<\mu<\overline{X}_n+\delta\right)$$

e quindi: $\quad 0.95 = P\left(\overline{X}_n-\delta<\mu<\overline{X}_n+\delta\right)$



che è proprio l'intervallo di confidenza per μ a livello di significatività del 95%.

1.1 Caso X distribuzione normale con varianza nota

Vediamo come calcolare effettivamente δ. Consideriamo prima il caso in cui la distribuzione di X sia normale e la *varianza sia nota*.

ESEMPIO: Si estrae un campione di numerosità 100 da una popolazione con distribuzione normale con varianza σ^2 = 225 nota e valore atteso incognito μ.
Vogliamo calcolare un intervallo di confidenza del valore atteso a livello di confidenza di $1-\alpha=0.95$ sapendo che la stima della media sul campione è $\overline{x}_n$ = 1450. Abbiamo visto che lo stimatore $\overline{X}_n$ ha valore atteso μ e varianza $\frac{\sigma^2}{n}=\frac{225}{100}=2.25$. Sappiamo, inoltre, che $\overline{X}_n$ ha ancora distribuzione normale: $\overline{X}_n \sim N(\mu, 2.25)$.
Vogliamo determinare δ tale che

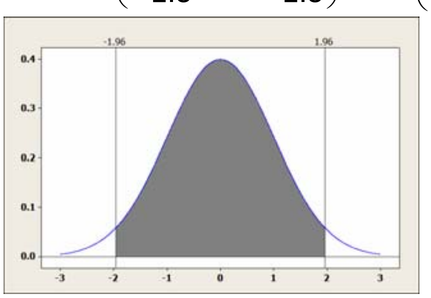
$$P\left(\mu-\delta<\overline{X}_n<\mu+\delta\right)=0.95$$

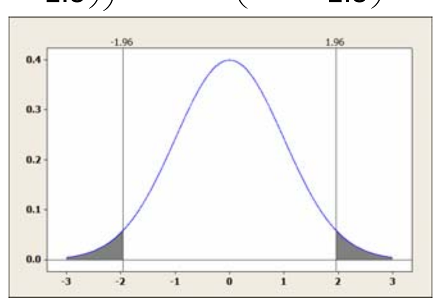
Per poter utilizzare le tavole della funzione di distribuzione cumulata di una variabile aleatoria Z normale (0,1), standardizziamo $\overline{X}_n$:

$$0.95 = P\left(\mu-\delta<\overline{X}_n<\mu+\delta\right)=P\left(\frac{\mu-\delta-\mu}{\sigma_{\overline{X}}}<\frac{\overline{X}_n-\mu}{\sigma_{\overline{X}}}<\frac{\mu+\delta-\mu}{\sigma_{\overline{X}}}\right)=P\left(-\frac{\delta}{1.5}<Z<\frac{\delta}{1.5}\right)$$

Siccome il grafico della densità di probabilità di Z è simmetrico rispetto all'asse verticale,
la probabilità che Z sia compresa fra i due valori $-\frac{\delta}{1.5}$ e $\frac{\delta}{1.5}$ è uguale a 1 meno la probabilità delle due parti esterne (le cosiddette "code"):

$$P\left(-\frac{\delta}{1.5}<Z<\frac{\delta}{1.5}\right)=1-\left(P\left(Z<-\frac{\delta}{1.5}\right)+P\left(Z>\frac{\delta}{1.5}\right)\right)=1-2P\left(Z<-\frac{\delta}{1.5}\right)$$





Quindi:

$$0.95 = 1 - 0.05 = 1 - 2P\left(Z < -\frac{\delta}{1.5}\right) \Leftrightarrow \ 0.05 = 2P\left(Z < -\frac{\delta}{1.5}\right) \Leftrightarrow \ 0.025 = P\left(Z < -\frac{\delta}{1.5}\right)$$

Dalle tavole si ottiene che $\frac{\delta}{1.5} = 1.96$, ossia δ=2.94, soddisfa le condizioni richieste.

Infine, sostituendo il valore numerico ottenuto sul campione, si ha che

$$(1450 - 2.94,\ 1450 + 2.94) = (1447.06,\ 1452.94)$$

è la *realizzazione* dell'intervallo di confidenza del valore atteso a livello 0.95.
Noi non sappiamo se il valore atteso di X nella popolazione apparenga o no effettivamente a questo intervallo.
Se avessimo avuto un'altra stima puntuale per la media, proveniente da un altro campione, avremmo avuto anche un diverso intervallo di confidenza.

Fra tutti i possibili intervalli di confidenza costruiti in questo modo sulla base di tutti i possibili campioni, il 95% contiene la media di X nella popolazione e il 5% non la contiene.

Riassumiamo i conti fatti per determinare un intervallo di confidenza a livello 1-α per la media di una variabile aleatoria con distribuzione normale di media μ incognita e varianza σ^2 nota:

1. Si utilizza come stimatore la media empirica $\overline{X}_n$ di un campione di numerosità n e si ricava la stima $\overline{x}_n$.
2. Si cerca sulle tavole della normale standardizzata, il valore z_α , tale che

$$P(Z < -z_\alpha) = 1 - \frac{\alpha}{2}.$$

3. Si costruisce l'intervallo aleatorio $\left(\overline{X}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \overline{X}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$;

   questo intervallo, che ha come estremi due variabili aleatorie, **ha probabilità 1-α** di contenere μ.
4. Si sostituisce il valore campionario $\overline{x}$ e si ottiene la realizzazione numerica dell'intervallo per il campione ottenuto. In formule:

$$I = \left(\overline{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \overline{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

Riportiamo nella seguente tabella i valori di z_α per alcuni α:

livello di conf.	90%	95%	99%
α	0.10	0.05	0.01
z_α	1.65	1.96	2.58

1.2 Caso X distribuzione normale con varianza sconosciuta

Quando la varianza della variabile aleatoria X è sconosciuta, si stima usando lo stimatore non distorto S^2.
La formula per calcolare l'intervallo di confidenza per il valore atteso è *leggermente* differente: non si usa z_α ma un altro valore che però è molto vicino a z_α se la numerosità campionaria è *molto grande* (maggiore di 100); in queste schede noi useremo l'approssimazione:

$$I = \left(\bar{x} - z_\alpha \frac{s}{\sqrt{n}}, \bar{x} + z_\alpha \frac{s}{\sqrt{n}}\right)$$

dove s è la realizzazione campionaria della standard deviation: $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_n\right)^2}$

1.3 Caso X con distribuzione qualsiasi e numerosità del campione grande

Cosa si può fare nel caso in cui la variabile X non abbia densità normale?

In alcuni casi è possibile calcolare in modo esplicito la distribuzione degli stimatori. Nella maggior parte dei casi, però, si utilizza l'approssimazione normale garantita dal Teorema del Limite Centrale. Abbiamo, infatti, visto che per n *sufficientemente* grande la media campionaria $\bar{X}_n$ ha *quasi* una distribuzione normale di media μ (pari a quella di X) e varianza σ^2/n. Quindi un intervallo di confidenza a livello $1-\alpha$ per la media di una variabile aleatoria con distribuzione NON normale di media μ incognita e varianza σ^2 nota sarà ancora

$$I = \left(\bar{x}_n - z_\alpha \frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_\alpha \frac{\sigma}{\sqrt{n}}\right)$$

con n, numerosità del campione, grande.
Resta da stabilire il significato di questa parola *grande.* Nella maggior parte dei casi una numerosità campionaria superiore a 30 è considerata accettabile per poter applicare il Teorema del Limite Centrale. Ricordiamo che i risultati sono approssimati e sono tanto più precisi quanto più alta è la numerosità campionaria.
Anche in questo caso, se la varianza non è nota si stima utilizzando lo stimatore non distorto S^2 e l'intervallo di confidenza è circa:

$$I = \left(\bar{x}_n - z_\alpha \frac{s}{\sqrt{n}}, \bar{x}_n + z_\alpha \frac{s}{\sqrt{n}}\right)$$

ESEMPIO:
Nel caso dei dati raccolti sul prezzo del latte, abbiamo:

- il prezzo medio campionario è $\bar{x}$ =1.34 euro
- la standar deviation campionaria è: s = 0.25 euro
- la numerosità campionaria è: 57

quindi $\frac{s}{\sqrt{n}}$ = 0.0333

Tutti questi valori sono forniti direttamente da Minitab; il valore di $\frac{s}{\sqrt{n}}$ è indicato nella colonna SE MEAN (cioè standard error della variabile aleatoria Media campionaria).

Variable	BENE	N	N*	**Mean**	**SE Mean**	**StDev**	Minimum	Q1	Median	Q3	Maximum
PREZZO	Latte	57	0	**1.3398**	**0.0333**	**0.2512**	0.6900	1.3000	1.3700	1.3900	2.5900

Non sappiamo se la variabile aleatoria che modella il prezzo di un litro di latte abbia distribuzione normale, ma essendo la numerosità campionaria maggiore di 30 possiamo usare il Teorema del limite centrale e trovare un intervallo di confidenza approssimato.
Se scegliamo α = 0.05, la realizzazione campionaria dell'intervallo di confidenza per il prezzo medio di un litro di latte è:

$$\left(\bar{x}_n - z_\alpha \frac{s}{\sqrt{n}}, \bar{x}_n + z_\alpha \frac{s}{\sqrt{n}}\right) = (1.3398 - 1.96 \times 0.0333 , 1.3398 + 1.96 \times 0.0333) = (1.28, 1.41)$$

Se scegliamo α = 0.01, la realizzazione campionaria dell'intervallo di confidenza per il prezzo medio di un litro di latte è:

$$\left(\bar{x}_n - z_\alpha \frac{s}{\sqrt{n}}, \bar{x}_n + z_\alpha \frac{s}{\sqrt{n}}\right) = (1.3398 - 2.58 \times 0.0333 , 1.3398 + 2.58 \times 0.0333) = (1.25, 1.43)$$

È meglio un intervallo di confidenza a livello di significatività del 95% o del 99%?
Sicuramente con un intervallo di confidenza a livello di significatività del 99% la probabilità di errore è più piccola rispetto a quella con un intervallo al 95%.
Ma nel primo caso l'ampiezza dell'intervallo è più grande: quello che si guadagna in precisione si perde in ampiezza.
Nell'esempio precedente:

- al 95% si ha δ = 6 centesimi di euro
- al 95% si ha δ = 9 centesimi di euro

ESERCIZIO
Calcolare la realizzazione campionaria di un intervallo di confidenza del prezzo medio degli altri beni raccolti

Variable	BENE	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
PREZZO	Benzina	64	0	1.2450	0.00382	0.0305	1.1200	1.2363	1.2450	1.2560	1.3000
	CD	35	0	1.069	0.108	0.637	0.290	0.800	0.950	1.000	3.900
	DVD	45	0	3.257	0.125	0.837	1.290	2.990	3.000	4.000	5.000
	Gasolio	54	0	1.1155	0.00557	0.0410	1.0120	1.0980	1.1120	1.1308	1.2500
	Olio	46	0	6.410	0.285	1.932	3.650	4.938	5.990	7.360	13.500

2. Intervalli di confidenza per la frequenza p

Nella scheda precedente abbiamo visto che uno stimatore per la frequenza di una variabile aleatoria dicotomica è dato da

$$\hat{P} = \frac{X_1 + \cdots + X_n}{n}$$

dove ciascuna delle variabili aleatorie X_1,..., X_n vale 1 (con probabilità p) oppure 0 (con probabilità 1-p) a seconda che si ottenga un successo o un insuccesso.
Abbiamo già visto che E($\hat{P}$)=p e Var($\hat{P}$)=$\frac{p(1-p)}{n}$.
Se abbiamo un campione di numerosità elevato possiamo approssimare la distribuzione di $\hat{P}$ con quella normale: $\hat{P} \sim N\left(p, \frac{p(1-p)}{n}\right)$. Anche la varianza è incognita perché dipendente ancora dal parametro p, ma si può stimare a partire dalla stima $\hat{p}$ del parametro p. Uno stimatore non distorto per Var($\hat{P}$) è

$$S^2_{\hat{P}} = \frac{\hat{P}(1-\hat{P})}{n-1}$$

Un intervallo di confidenza per p a livello di significatività 1-α è quindi

$$I=\left(\hat{p}-z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}},\hat{p}+z_{\alpha}\sqrt{\frac{\hat{p}(1-\hat{p})}{n-1}}\right).$$

ESEMPIO: Una popolazione di animali è composta da una razza con il pelo uniforme e da una con il pelo striato. Si osservano 100 animali e si nota che 70 di questi hanno il pelo striato. Vogliamo calcolare un intervallo di confidenza a livello del 99% per la popolazione di animali dal pelo striato. Utilizziamo le formule precedenti scegliendo

$$n=100, \qquad \hat{p}=0.70, \qquad \alpha=0.01, \qquad z_{\alpha}=2.58.$$

Sostituendo otteniamo che la realizzazione dell'intervallo di confidenza per p è:

$$I=\left(0.70 - 2.58\sqrt{\frac{0.7x0.3}{99}}, 0.70+2.58\sqrt{\frac{0.7x0.3}{99}}\right)=(0.58, 0.82).$$

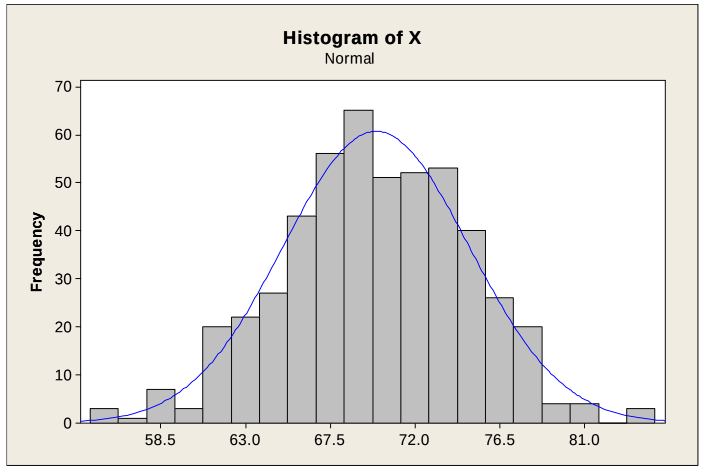
ESERCIZIO SU: Campionamento – Teorema del limite centrale – Intervalli di confidenza

Si vuole stimare la media di una grandezza in una popolazione di 500 unità. Si modella il fenomeno con una variabile aleatoria X.

	1	2	3	4	5	6	7	8	9	10
0	70.3	65.8	63.0	62.0	69.0	70.8	73.8	70.5	72.1	75.8
1	78.4	66.0	69.2	62.1	77.0	70.6	68.9	65.9	76.1	61.3
2	73.2	70.6	75.2	72.5	70.0	68.6	66.8	69.1	73.8	75.6
3	68.4	68.2	62.3	76.9	72.2	67.4	71.2	77.0	66.2	72.0
4	67.4	69.7	68.2	62.7	75.1	63.2	72.9	62.8	69.6	62.1
5	68.2	66.4	72.4	65.2	67.3	68.6	74.2	66.7	67.9	68.6
6	66.2	72.5	65.7	69.7	75.4	78.2	73.8	69.2	69.4	78.3
7	72.1	75.6	76.3	71.7	75.9	78.6	69.5	72.2	63.3	67.4
8	60.2	74.1	74.8	63.5	66.2	67.8	65.5	68.1	74.0	68.9
9	69.2	61.8	64.5	61.3	71.0	71.4	72.5	71.7	72.9	69.6
10	62.7	71.1	77.6	73.5	66.8	72.0	67.5	69.2	69.8	77.9
11	76.8	80.5	68.6	67.2	63.4	70.5	61.8	68.9	64.4	72.7
12	72.9	66.6	58.5	73.3	73.7	67.5	64.1	70.1	76.8	77.3
13	69.9	68.7	75.3	69.6	67.6	66.7	68.5	77.4	75.6	67.5
14	71.2	68.8	65.9	70.0	65.3	67.0	54.9	71.0	63.1	77.8
15	71.9	67.4	74.6	65.0	72.1	70.6	72.5	74.3	75.8	62.0
16	65.8	74.5	62.6	76.6	75.3	78.0	71.6	77.8	64.4	71.7
17	66.7	71.0	69.5	68.3	72.9	70.6	74.5	71.5	64.4	78.5
18	65.0	73.6	71.2	70.9	69.4	75.9	83.6	73.5	73.3	68.6
19	73.1	75.5	76.1	69.7	65.3	72.7	72.4	65.8	69.2	61.1
20	69.0	68.7	61.2	71.3	65.9	75.3	68.2	67.6	72.6	69.2
21	75.6	75.7	68.1	73.3	74.5	72.0	80.1	66.5	64.0	70.3
22	68.2	70.4	69.5	72.5	69.3	72.8	73.2	72.0	70.9	73.6
23	84.5	73.4	68.2	71.2	73.7	66.2	74.0	75.6	74.1	65.1
24	61.1	77.2	75.0	73.1	71.6	68.1	69.8	63.5	69.5	84.7
25	63.6	72.7	76.5	70.7	72.5	63.5	67.4	67.0	74.6	66.7
26	66.6	68.3	64.8	75.5	75.8	68.6	70.2	68.0	70.0	72.1
27	71.3	69.7	64.7	64.3	66.1	74.1	72.9	65.6	71.2	64.0
28	68.7	58.2	69.7	70.0	68.4	69.6	74.8	70.2	72.7	73.6
29	73.3	62.7	62.3	72.0	72.8	73.4	79.5	65.4	74.6	63.5
30	73.3	67.8	68.2	80.2	65.2	63.3	70.1	69.3	66.9	71.1
31	75.0	80.8	66.8	74.5	69.4	67.5	68.8	75.6	62.0	73.5
32	74.8	75.1	72.7	70.1	77.4	72.8	74.4	61.9	55.6	73.3
33	70.4	66.4	70.6	75.2	71.7	67.8	58.4	68.8	66.9	64.2
34	71.4	66.2	78.5	64.0	59.4	60.6	66.6	68.3	70.2	61.1
35	71.3	73.4	76.8	55.6	74.3	70.9	77.8	69.1	65.7	69.0
36	67.9	64.2	73.5	71.2	58.7	71.7	68.8	76.2	68.8	67.5
37	77.8	65.8	69.4	67.8	71.3	67.0	66.5	78.4	66.6	64.9
38	72.2	72.8	66.2	64.3	68.4	68.2	65.0	65.3	73.7	62.1
39	73.2	75.2	67.5	73.7	67.3	73.7	62.5	66.8	69.8	69.9
40	63.1	72.2	75.2	60.9	74.6	69.6	66.0	67.7	61.8	80.2
41	75.3	69.4	69.2	68.2	76.9	77.1	73.1	68.6	78.3	70.1
42	70.5	68.0	66.3	76.2	65.4	67.9	63.8	73.5	70.0	70.7
43	72.6	72.0	75.5	72.2	68.2	73.8	77.0	73.1	63.9	71.5
44	66.4	73.0	70.4	76.1	76.9	69.4	68.6	68.9	64.5	70.8
45	67.7	80.9	59.2	62.8	73.2	71.4	58.7	64.7	74.6	67.4
46	62.7	69.0	66.7	70.6	68.7	67.9	76.7	73.4	70.2	67.1
47	75.6	68.9	68.8	71.9	66.2	70.2	72.3	66.2	67.5	73.9
48	67.3	61.9	72.3	71.5	78.7	81.3	62.1	59.1	66.9	72.0
49	68.1	66.6	75.0	76.0	57.4	61.2	73.3	77.9	64.2	64.8

Si vuole stimare la media μ_X della variabile X nella popolazione, sapendo che la standard deviation di X è: $\mathrm{std}(X) = \mathbf{4.93202}$

1) Ciascuno studente estragga dalla popolazione 5 campioni casuali semplici di numerosità 20 utilizzando i numeri casuali riportati nella pagina seguente. (Campione casuale semplice: estratto con ripetizione da distribuzione uniforme)



La variabile X nella popolazione ha una distribuzione “a campana”; quindi per campioni di 20 unità sperimentali la distribuzione della variabile aleatoria $\bar{X}_{20}$ può essere approssimata con quella di una variabile aleatoria normale.

(Nota: in genere si effettua tale approssimazione per numerosità maggiori di 30).

Unità sperim.	Campione 1	Campione 2	Campione 3	Campione 4	Campione 5
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					

2) Scrivere la formula per l’intervallo di confidenza per μ_X a livello di significatività del 95%, quando la varianza è nota.

3) Per ciascun campione estratto calcolare la media campionaria e l'intervallo di confidenza al 95% per μ_X. Usare 5 cifre dopo la virgola.

	$\overline{X}_{20}$	semiampiezza intervallo	limite sinistro	limite destro
1				
2				
3				
4				
5				

4) Quando tutti gli studenti hanno terminato verrà fornito il valore vero di μ_X e si verificherà quanti intervalli calcolati con la formula usuale contengono effettivamente μ_X. Quanti intervalli si prevede contengano μ_X?

Per campionare. Qui sotto, per ciascuno studente, sono fornite 5 colonne con 20 numeri casuali fra 1 e 500 (numerosità della popolazione) estratti con ripetizione. Come utilizzare questi numeri.

Il primo numero casuale del primo studente 396. Quindi deve considerare il valore di X per la 396-esima unità sperimentale. Quale valore assume la variabile X per 396-esima unità campionaria? È scritto nella tabella dei dati della popolazione nella riga 39-esima e nella colonna 6: è 73.7

Se il numero è di una sola cifra, ad esempio 4, si cerca il valore presente alla riga 0 e alla colonna 4, cioè 62.0.

Se il numero è di due cifre, del tipo 40, si cerca il valore alla riga 30 e alla colonna 10

1

396	342	433	247	301
52	213	351	193	54
493	282	498	278	164
24	331	424	493	474
214	213	425	479	251
377	329	282	26	330
390	354	367	378	51
170	256	115	120	405
468	136	197	488	169
4	150	418	313	239
425	388	1	350	467
483	485	191	269	386
261	87	107	443	204
469	14	118	162	265
111	463	395	132	186
250	266	480	192	445
308	389	38	203	396
241	313	370	282	82
324	21	497	174	255
441	360	455	379	100

2

133	206	63	145	330
494	200	218	473	307
66	417	468	332	57
254	417	299	435	452
441	217	231	237	453
71	379	94	303	482
240	234	60	427	335
151	225	298	288	67
355	271	214	331	251
19	232	78	60	82
476	50	100	246	403
253	170	202	59	27
383	396	335	341	181
257	164	27	373	82
28	250	112	209	20
133	474	400	286	470
355	147	38	388	240
411	396	145	244	186
90	322	417	354	27
2	402	287	457	391

3

57	348	226	315	411
230	395	451	72	337
92	455	385	154	241
210	432	317	119	197
474	286	377	100	315
499	142	214	461	192
490	120	440	482	103
393	176	483	392	59
250	91	131	91	159
80	444	80	458	214
494	166	273	461	390
407	383	339	260	434
385	106	310	221	474
263	119	472	369	360
63	258	383	300	72
112	260	287	28	316
272	239	365	58	258
288	289	171	302	120
393	98	491	399	54
111	315	222	475	295

4

195	385	233	29	406
23	365	237	291	117
272	490	456	371	27
304	493	485	233	6
243	359	278	116	282
187	14	397	60	465
88	463	24	114	430
205	301	101	250	237
22	160	292	8	40
487	68	227	158	3
243	225	465	451	479
481	25	220	332	38
234	286	89	457	187
330	279	261	331	290
194	456	21	353	217
388	213	94	53	483
475	98	472	246	320
243	187	397	211	391
79	393	168	5	120
329	432	441	141	403

5

110	106	340	50	111
492	377	459	392	59
245	95	410	230	172
234	185	467	237	58
250	235	120	256	315
451	414	102	153	34
402	378	268	81	279
247	150	315	229	46
137	212	106	139	150
103	270	137	107	81
234	173	302	335	291
228	127	405	368	219
6	303	400	223	182
472	355	165	98	252
350	194	62	286	23
105	143	166	187	276
149	402	243	431	369
291	417	138	336	97
409	177	136	494	128
193	489	419	10	24

6

154	42	267	93	307
241	345	118	37	255
362	283	68	158	376
365	415	136	438	350
259	180	384	54	483
183	232	347	202	93
335	57	197	239	261
212	198	252	467	388
474	74	186	41	499
221	171	304	15	243
368	329	126	318	20
9	315	439	75	467
40	115	426	493	167
62	67	34	440	69
352	34	335	450	227
307	54	151	132	469
457	38	101	161	480
304	139	46	65	353
24	406	474	406	111
207	74	342	498	213

7

20	244	387	12	385
376	319	185	389	242
257	363	81	349	465
264	64	314	478	64
313	248	247	126	84
134	348	168	490	188
23	52	147	77	167
480	376	73	469	36
471	483	454	413	473
116	493	93	91	210
130	434	149	69	344
183	269	10	297	253
9	358	383	434	169
72	324	301	92	49
30	14	305	332	401
315	494	182	219	333
212	96	289	374	135
104	319	338	371	242
380	164	96	65	56
362	451	123	211	100

8

154	205	325	489	328
198	320	432	139	210
406	234	454	183	426
421	186	374	97	275
192	105	357	357	463
82	272	226	381	228
436	52	45	478	62
478	284	213	187	307
328	26	316	171	221
333	421	109	491	500
5	333	48	185	461
94	473	296	407	134
216	453	59	120	478
296	199	121	414	486
126	293	48	207	409
247	4	123	221	361
427	294	112	48	476
260	287	350	468	129
440	30	326	116	264
47	493	395	61	330

9

352	80	131	149	488
390	397	420	188	390
308	149	169	465	239
372	33	327	27	245
232	183	60	27	374
134	47	183	211	410
184	91	90	70	316
234	476	439	116	47
470	465	459	297	436
478	408	209	264	473
12	470	490	499	183
109	300	411	426	374
211	483	158	91	189
250	307	427	357	445
289	376	118	296	477
171	333	29	138	406
219	256	32	334	413
8	58	275	489	442
9	318	469	60	291
129	44	83	325	494

10

84	269	447	169	49
169	245	25	190	149
480	217	101	351	176
62	263	184	229	450
354	46	30	247	29
181	215	40	164	273
451	66	142	277	454
164	339	298	488	346
1	448	151	320	365
58	432	118	73	106
240	90	67	253	244
478	236	247	334	11
134	371	24	220	266
91	233	83	235	170
105	260	20	261	286
267	421	262	479	32
458	457	246	6	239
396	112	32	92	85
32	156	273	136	269
440	195	363	76	15

11

453	477	484	85	191
143	145	152	441	472
230	399	495	230	436
278	138	443	419	42
97	293	301	26	171
153	331	203	424	398
244	377	218	448	318
467	231	66	164	113
92	321	272	127	466
102	500	32	496	399
436	1	175	314	14
202	186	449	146	326
7	386	106	253	279
438	376	421	443	430
346	378	307	225	44
59	421	93	41	193
139	116	145	163	231
309	409	82	388	450
14	132	368	38	218
143	393	296	445	337

12

448	376	214	332	392
421	333	310	368	329
472	40	379	408	142
111	173	126	448	347
56	80	378	190	209
160	435	279	106	104
445	70	219	344	166
313	330	281	171	322
321	221	475	367	473
151	455	128	394	349
401	428	224	28	415
188	195	281	427	466
78	72	253	160	180
436	261	128	476	346
410	324	226	171	158
97	193	460	33	263
496	277	478	466	177
353	188	369	127	18
203	7	290	83	430
163	187	25	120	164

13

330	213	441	236	489
339	55	404	116	233
8	403	453	132	467
64	362	214	287	202
373	416	54	1	258
381	221	134	62	110
242	425	447	111	486
353	432	483	409	297
480	158	414	211	74
389	369	212	201	39
263	334	253	226	330
112	211	320	399	247
87	414	287	78	481
450	15	76	265	57
358	7	416	186	171
333	37	49	225	464
238	77	96	369	396
394	312	246	458	276
273	411	296	77	44
383	66	93	362	426

14

296	416	3	194	329
220	235	78	11	285
419	295	242	365	490
429	53	304	406	104
70	142	262	246	113
127	492	31	265	276
135	88	283	261	487
251	55	306	240	125
456	295	324	286	177
222	131	216	470	481
358	347	453	357	81
163	313	91	65	447
12	321	26	25	113
52	13	118	396	209
444	383	458	369	300
356	335	58	11	483
447	497	156	128	50
47	119	334	67	162
313	351	362	307	296
314	106	133	229	474

15

491	318	80	285	309
264	295	253	120	415
171	351	254	361	383
133	122	94	477	36
330	283	174	344	233
55	148	392	262	2
169	364	428	387	105
222	407	163	202	357
401	340	2	242	453
35	377	344	325	23
326	335	115	114	433
276	182	335	328	5
308	142	55	317	167
96	341	392	342	433
419	363	103	74	107
392	146	384	115	229
177	258	129	84	221
481	131	378	344	415
410	137	399	332	443
244	18	451	446	47

16

396	403	170	344	387
327	384	40	451	124
152	242	153	6	280
197	301	94	376	187
131	258	25	482	298
401	394	334	197	334
454	282	311	155	274
409	416	98	129	448
461	147	456	152	393
36	266	133	162	455
402	34	458	456	275
273	181	336	349	423
72	192	394	380	469
323	85	435	227	493
64	135	116	43	339
269	47	390	173	284
142	461	40	227	236
214	113	103	376	19
215	108	429	466	414
75	251	409	86	409

17

207	484	395	412	179
327	161	489	12	398
495	344	257	185	72
129	101	319	452	395
496	206	235	256	41
269	423	337	313	329
27	168	162	255	95
148	113	393	394	242
142	302	52	395	433
81	169	149	159	195
418	209	194	500	264
27	374	348	420	56
117	317	272	334	67
429	10	181	354	54
63	374	117	428	432
179	358	405	497	186
354	329	412	149	88
134	370	415	384	333
444	488	335	255	57
455	262	369	459	123

18

417	172	185	289	177
99	99	434	189	128
49	133	274	302	18
400	184	130	490	343
24	80	115	187	493
119	203	351	210	245
346	425	339	27	149
29	111	282	166	209
87	498	470	393	150
178	240	27	290	493
354	62	199	149	53
9	386	242	314	246
226	398	486	377	211
443	168	342	437	11
53	451	282	221	469
256	367	16	232	430
414	479	99	82	307
7	76	107	130	322
1	161	423	251	341
66	478	47	444	476

19

50	179	197	498	80
477	372	352	197	88
387	63	453	396	410
59	150	319	153	393
337	339	437	164	195
335	76	495	373	307
449	1	77	216	2
464	336	1	59	240
398	29	334	471	89
334	428	307	342	495
142	341	85	449	226
72	235	420	481	255
299	376	435	372	418
444	145	268	406	52
165	391	134	271	35
374	255	433	468	41
163	1	375	213	492
482	115	31	182	301
277	341	123	312	304
422	444	477	450	222

20

56	139	330	198	329
159	417	172	252	455
107	269	399	16	39
294	366	116	215	352
498	383	412	46	312
317	271	342	151	340
177	470	153	149	429
307	80	492	150	382
256	453	142	91	381
247	350	62	135	112
34	493	435	44	380
91	261	29	132	180
292	134	397	8	186
315	478	40	201	154
96	226	338	126	256
494	409	184	103	490
312	446	137	241	123
388	405	150	272	227
68	429	116	286	172
143	476	12	459	323

21

109	445	199	110	306
41	407	295	225	460
298	66	474	65	192
322	110	131	129	36
311	334	367	24	438
280	45	124	381	494
363	479	55	223	56
209	459	149	104	462
455	413	116	87	402
500	95	8	373	222
173	398	315	411	397
388	439	208	483	318
102	497	413	368	130
97	121	141	390	409
464	269	335	190	187
463	325	473	70	99
232	316	281	349	327
359	392	424	209	340
137	153	314	227	380
464	26	384	365	399

22

22	415	244	16	409
482	79	39	161	331
19	400	380	11	337
375	55	311	341	36
46	401	35	70	463
325	51	262	477	32
126	208	388	124	169
306	364	303	454	373
475	358	79	368	301
140	347	52	311	25
287	187	89	469	364
243	225	334	90	49
397	268	85	254	409
72	64	376	303	112
358	432	54	149	439
208	334	377	329	461
336	329	135	237	265
480	244	307	98	143
240	207	314	84	400
64	160	105	57	199

23

309	285	252	468	345
364	165	413	390	152
237	365	16	181	306
416	366	479	249	118
341	43	254	267	64
330	424	157	149	177
87	457	242	461	489
439	175	106	205	131
363	313	396	479	489
114	379	279	180	128
106	54	486	38	158
418	289	404	460	337
260	252	489	249	12
262	80	122	41	411
20	416	24	390	406
66	9	468	23	269
170	151	347	223	138
428	277	256	258	319
336	170	479	159	390
97	494	451	241	281

24

259	44	344	126	353
109	452	466	425	7
240	371	212	19	260
160	295	46	463	359
228	388	356	305	176
267	62	343	393	350
165	406	381	51	275
174	261	139	369	380
340	136	238	94	168
366	180	153	457	411
278	220	434	178	40
152	174	377	420	124
169	5	491	461	432
494	184	296	409	254
376	360	455	109	425
332	448	10	115	490
156	345	74	137	450
323	250	487	108	269
58	104	18	302	346
68	67	213	16	223

25

91	344	167	282	403
473	266	477	311	390
415	338	341	303	392
295	185	405	241	249
143	396	322	382	92
97	282	128	472	461
439	118	317	417	150
242	298	236	382	365
115	347	106	60	162
118	44	300	350	127
199	353	14	79	292
184	404	29	277	259
173	109	34	42	155
375	438	130	246	190
403	67	386	355	484
110	39	411	96	77
390	406	229	238	491
42	30	39	399	272
414	479	25	245	29
197	201	352	473	75

ESERCIZI

1) Da 400 lanci di una moneta sono risultati 175 esiti “testa” e 225 esiti “croce”.

a) Trovare un intervallo di confidenza al 90% per la probabilità di esito “testa”.
b) Trovare un intervallo di confidenza al 99% per la probabilità di esito “testa”.
c) Questa moneta sembra truccata? Giustificare la risposta.

2) Spiegare, eventualmente con un esempio, perché l’intervallo di confidenza di un parametro può non contenere il parametro che si vuole stimare.

3) Si vogliono effettuare stime per la quantità di sostanza attiva in una unità di un certo farmaco (espressa in mg). Si può ipotizzare che la variabile casuale X che rappresenta la quantità di sostanza attiva abbia distribuzione normale. A tal fine si effettua un campionamento casuale di 100 unità del farmaco. Per questo campione si ottiene:

$$\sum_{i=1}^{n} x_i = 222.91 \quad \text{e} \quad \sum_{i=1}^{n} x_i^2 = 1154.8$$

Calcolare una stima puntuale e un intervallo di confidenza a livello di significatività del 99% per la media di sostanza attiva del farmaco.

4) Dai dati del censimento del 1991 risulta che il numero di abitazioni di una città è 300 000 e che la media dell’epoca di costruzione delle abitazioni è 1815 e lo scarto quadratico medio è 50 anni.
Uno statistico calcola l’intervallo di confidenza per la media dell’epoca di costruzione al 95%. Commentare.

5) Si determina l’ampiezza 2δ di un intervallo di confidenza a livello fissato $1\text{-}\alpha$ per la media di una variabile aleatoria normale di varianza nota, sulla base di un campione di numerosità n. Quanto numeroso deve essere il campione se si vuole che l’intervallo risultante, con lo allo stesso livello, abbia ampiezza pari ad un terzo di quello che si ottiene con un campione di numerosità n?

6) Sia X una variabili aleatorie di Bernoulli di parametro p, siano $X_1,\dots,X_n$ le variabili aleatorie campionarie e sia $\hat{P}$ lo stimatore di p.

a) Scrivere (*in funzione di* n, p *)* la semiampiezza δ dell’intervallo di confidenza per p a livello di significatività del 95%.
b) Per quale valore di p la semiampiezza δ è massima?
c) Come deve essere scelto n affinché la semiampiezza δ sia minore o uguale a 0.05?

7) Sia X una variabile aleatoria con distribuzione normale di media μ e varianza σ^2 entrambe sconosciute. Per stimare il parametro μ si effettua un campionamento di numerosità 16. Si indichi con I_{16}^{α} la realizzazione campionaria dell’intervallo di confidenza per μ a livello di significatività fissato $1-\alpha$.
Si amplia il campione percedente di altre 9 unità (ottenendo un campione totale di 25 elementi); si indichi con I_{25}^{α} la realizzazione campionaria dell’intervallo di confidenza per μ nel campione totale allo stesso livello di significatività.
Dire se le seguenti relazioni sono vere, false o se non si può affermare né una cosa né l’altra:

a) $I_{16}^{\alpha} \subset I_{25}^{\alpha}$ b) $I_{16}^{\alpha} \supset I_{25}^{\alpha}$ c) $I_{16}^{\alpha} \cap I_{25}^{\alpha} = \varnothing$ d) $I_{16}^{\alpha} \cap I_{25}^{\alpha} \neq \varnothing$

8) Sia X una variabile aleatoria con distribuzione normale di media sconosciuta e varianza nota. Indichiamo con (A, B) l’intervallo di confidenza per la media calcolato su un campione di n elementi. È vero che A e B sono variabili aleatorie?

9) Sia X una variabile aleatoria con distribuzione normale di media μ e varianza σ^2 entrambe sconosciute. Sulla base di un campione di numerosità n si calcola un intervallo di confidenza per μ al livello del 95%. Esite un intervallo di confidenza per μ, allo stesso livello, su un campione di uguale numerosità con ampiezza minore del precedente?

10) Uno scienziato sostiene che il 9% delle stelle ammette un sistema planetario.

a) Determinare la probabilità che su 1000 stelle almeno 100 abbiano un sistema planetario, secondo le ipotesi dello scienziato.

b) Sulle 80 stelle più vicine alla terra se ne sono trovate 3 con un sistema planetario. Si calcoli un intervallo di confidenza a livello del 5% per la frequenza relativa delle stelle vicine alla terra con sistema planetario.

11) A parità di altre condizioni (numerosità campionaria, ...) è vero che l'ampiezza dell'intervallo di confidenza per il valore atteso è tanto maggiore quanto è minore il livello 1-α? Giustificare la risposta.

12) Sia X una variabile aleatoria di legge normale. Si effettua un campionamento di numerosità 10 e si ottengono i seguenti valori campionari:

24.2 22.5 26.7 27.0 28.2 21.3 23.8 24.5 23.2 22.9

a) Calcolare un intervallo di confidenza per la media a livello di significatività 0.90

b) Supponendo che la varianza sia nota e pari a 2, indicare la minima numerosità campionaria affinché l'ampiezza dell'intervallo di confidenza sia minore o uguale a 1, mantenendo lo stesso livello di significatività.