Curva Gaussiana – 1

valori di u	Nell'intervallo $[\mu - u\sigma, \mu + u\sigma]$	Fuori dell'intervallo $[\mu - u\sigma, \mu + u\sigma]$	Nell'intervallo $[\mu + u\sigma, +\infty)$
0	0	1	$0,5$
$0,2$	$0,1586$	$0,8414$	$0,4207$
$0,4$	$0,3108$	$0,6892$	$0,3446$
$0,6$	$0,4514$	$0,5486$	$0,2743$
$0,8$	$0,5762$	$0,4238$	$0,2119$
1	$0,6826$	$0,3174$	$0,1587$
$1,2$	$0,7698$	$0,2302$	$0,1151$
$1,4$	$0,8384$	$0,1616$	$0,0808$
$1,6$	$0,8904$	$0,1096$	$0,0548$
$1,8$	$0,9282$	$0,0718$	$0,0359$
2	$0,9544$	$0,0456$	$0,0228$
$2,2$	$0,9722$	$0,0278$	$0,0139$
$2,4$	$0,9836$	$0,0164$	$0,0082$
$2,6$	$0,9906$	$0,0094$	$0,0047$
$2,8$	$0,9950$	$0,0050$	$0,0025$
3	$0,9974$	$0,0026$	$0,0013$
$3,2$	$0,9986$	$0,0014$	$0,0007$

Curva Gaussiana – 2

aree sottese dalla curva gaussiana sull'intervallo [μ , $\mu + z\,\sigma$]

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,10	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,20	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,30	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,40	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,50	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,60	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,70	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,80	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,90	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,00	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,10	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,20	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015
1,30	0,4032	0,4049	0,4066	0,4082	0,4099	0,4115	0,4131	0,4147	0,4162	0,4177
1,40	0,4192	0,4207	0,4222	0,4236	0,4251	0,4265	0,4279	0,4292	0,4306	0,4319
1,50	0,4332	0,4345	0,4357	0,4370	0,4382	0,4394	0,4406	0,4418	0,4429	0,4441
1,60	0,4452	0,4463	0,4474	0,4484	0,4495	0,4505	0,4515	0,4525	0,4535	0,4545
1,70	0,4554	0,4564	0,4573	0,4582	0,4591	0,4599	0,4608	0,4616	0,4625	0,4633
1,80	0,4641	0,4649	0,4656	0,4664	0,4671	0,4678	0,4686	0,4693	0,4699	0,4706
1,90	0,4713	0,4719	0,4726	0,4732	0,4738	0,4744	**0,4750**	0,4756	0,4761	0,4767
2,00	0,4772	0,4778	0,4783	0,4788	0,4793	0,4798	0,4803	0,4808	0,4812	0,4817
2,10	0,4821	0,4826	0,4830	0,4834	0,4838	0,4842	0,4846	0,4850	0,4854	0,4857
2,20	0,4861	0,4864	0,4868	0,4871	0,4875	0,4878	0,4881	0,4884	0,4887	0,4890
2,30	0,4893	0,4896	0,4898	0,4901	0,4904	0,4906	0,4909	0,4911	0,4913	0,4916
2,40	0,4918	0,4920	0,4922	0,4925	0,4927	0,4929	0,4931	0,4932	0,4934	0,4936
2,50	0,4938	0,4940	0,4941	0,4943	0,4945	0,4946	0,4948	0,4949	**0,4951**	0,4952
2,60	0,4953	0,4955	0,4956	0,4957	0,4959	0,4960	0,4961	0,4962	0,4963	0,4964
2,70	0,4965	0,4966	0,4967	0,4968	0,4969	0,4970	0,4971	0,4972	0,4973	0,4974
2,80	0,4974	0,4975	0,4976	0,4977	0,4977	0,4978	0,4979	0,4979	0,4980	0,4981
2,90	0,4981	0,4982	0,4982	0,4983	0,4984	0,4984	0,4985	0,4985	0,4986	0,4986
3,00	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990

Teorema del Limite Centrale

Problema. Determinare come la media campionaria $\overline{x}$ e la deviazione standard campionaria s misurano la media μ e la deviazione standard σ della popolazione.

Sia data una popolazione numerica di media μ e deviazione standard σ e si estraggano da essa dei campioni casuali $\mathcal{C}_1$, $\mathcal{C}_2$, . . . , $\mathcal{C}_M$, ciascuno formato da n individui, con $n > 30$. Possiamo calcolare la media campionaria $\overline{x}_i$ di ciascun campione $\mathcal{C}_i$ $(i = 1, 2, \ldots, M)$ ed ottenenere così un nuovo insieme numerico, quello delle M medie campionarie.

Come si distribuiscono le medie campionarie?

Manifestano una tendenza in un certo senso *universale*, seguendo una *legge generale*, oppure il loro comportamento dipende dalla distribuzione della popolazione?

Teorema del Limite Centrale

Teorema. Sia data una popolazione numerica infinita di media μ e deviazione standard σ da cui vengono estratti dei campioni casuali formati ciascuno da n individui, con n abbastanza grande ($n > 30$). La distribuzione delle medie campionarie tende a una distribuzione gaussiana di media $\mu_{\overline{x}} = \mu$ e deviazione standard $\sigma_{\overline{x}} = \dfrac{\sigma}{\sqrt{n}}$.

In altre parole, anche in una popolazione che non segue il modello gaussiano, le medie campionarie, se calcolate su campioni abbastanza grandi, tendono a distribuirsi secondo una legge gaussiana.

Intervalli di Confidenza – 1

Come si utilizza il teorema del limite centrale?

Supponiamo di avere un campione casuale abbastanza grande.
Calcoliamo la *media campionaria* $\overline{x}$.

La distribuzione delle medie campionarie è gaussiana, quindi:

- il 99% dei dati cade nell'intervallo $[\mu - 2.58\,\sigma_{\overline{x}}\,,\ \mu + 2.58\,\sigma_{\overline{x}}]$, cioè per il 99% dei campioni:

$$\mu - 2.58\,\sigma_{\overline{x}} \le \overline{x} \le \mu + 2.58\,\sigma_{\overline{x}}$$

- il 95% dei dati cade nell'intervallo $[\mu - 1.96\,\sigma_{\overline{x}}\,,\ \mu + 1.96\,\sigma_{\overline{x}}]$, cioè per il 95% dei campioni:

$$\mu - 1.96\,\sigma_{\overline{x}} \le \overline{x} \le \mu + 1.96\,\sigma_{\overline{x}}$$

- ...

Intervalli di Confidenza – 2

Lette in termini di μ, le disuguaglianze precedenti definiscono gli intervalli di confidenza per la media μ della popolazione:

- intervallo di confidenza al 99%: $\overline{x} - 2.58\,\sigma_{\overline{x}} \leq \mu \leq \overline{x} + 2.58\,\sigma_{\overline{x}}$
- intervallo di confidenza al 95%: $\overline{x} - 1.96\,\sigma_{\overline{x}} \leq \mu \leq \overline{x} + 1.96\,\sigma_{\overline{x}}$
- ...

L'ampiezza degli intervalli di confidenza è espressa in funzione di

$$\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{n}}$$

che dipende dalla deviazione standard, incognita, della popolazione.

Intervalli di Confidenza – 3

Si può dimostrare che la *deviazione standard campionaria*

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

approssima bene la deviazione standard σ della popolazione.

Quindi, possiamo scrivere gli intervalli di confidenza nella forma:

- al 99%, $\overline{x} - 2.58 \frac{s}{\sqrt{n}} \le \mu \le \overline{x} + 2.58 \frac{s}{\sqrt{n}}$
- al 95%, $\overline{x} - 1.96 \frac{s}{\sqrt{n}} \le \mu \le \overline{x} + 1.96 \frac{s}{\sqrt{n}}$
- ...

che dipende solo dai dati campionari ($\overline{x}$, s, n).

Esercizio 1. Si vuole stimare l'età media degli utenti di una biblioteca civica. A questo scopo si seleziona un campione casuale composto da $n = 100$ persone avente media $\overline{x} = 29$ anni e deviazione standard $s = 8$ anni. Trovare intervalli di confidenza per l'età media μ al 95% ed al 99%.

Esercizi

Soluzione: poiché il campione è composto da $n = 100 > 30$ individui, possiamo applicare il teorema del limite centrale.

- Nel 95% dei casi la media μ appartiene all'intervallo
$$29 - 1.96 \cdot \frac{s}{\sqrt{n}} \leq \mu \leq 29 + 1.96 \cdot \frac{s}{\sqrt{n}}$$
Inserendo i dati, concludiamo che $27.43 \leq \mu \leq 30.57$ con un grado di fiducia pari al 95%.

- Nel 99% dei casi la media μ appartiene all'intervallo
$$29 - 2.58 \cdot \frac{s}{\sqrt{n}} \leq \mu \leq 29 + 2.58 \cdot \frac{s}{\sqrt{n}}$$
Inserendo i dati, concludiamo che $26.93 \leq \mu \leq 31.07$ con un grado di fiducia pari al 99%.

Esercizio 2. Nell'esercizio precedente si supponga che i dati $\overline{x} = 29$ anni e deviazione standard $s = 8$ anni siano stati ottenuti da un campione casuale composto da $n = 400$ persone. Trovare i nuovi intervalli di confidenza per l'età media μ al 95% ed al 99%.

Soluzione: l'unico cambiamento riguarda il fatto che

$$\frac{s}{\sqrt{n}} = \frac{8}{20} = \frac{2}{5}.$$

L'intervallo di confidenza al 95% è $28.21 \leq \mu \leq 29.79$.

L'intervallo di confidenza al 99% è $27.96 \leq \mu \leq 30.04$.

Rispetto all'esercizio precedente gli intervalli si sono ridotti di ampiezza, dunque la stima è più precisa. Il maggior grado di precisione è dovuto al fatto che i dati provengono da un campione più ampio.

Esercizio 3. Il diametro di certe biglie di acciaio segue una distribuzione gaussiana di media $\mu = 6.2$mm e deviazione standard $\sigma = 0.05$mm. Dire quale è la percentuale di biglie con diametro compreso tra 6.3mm e 6.35mm.

Soluzione: $[6.3\,,\,6.35] = [\mu + 2\sigma, \mu + 3\sigma] \;\Rightarrow\; 2.15\%$

Esercizio 4. Si vuole stimare l'età media μ di una popolazione di pazienti affetti da una certa malattia. Su un campione casuale composto da 576 pazienti affetti dalla malattia risulta un'età media $\bar{x} = 12$ anni e una deviazione standard campionaria $s = 4$ anni. Trovare l'intervallo di confidenza al 95% per l'età media μ dei malati.

Soluzione: $$\left[12 - 1.96 \cdot \frac{4}{\sqrt{576}}, 12 + 1.96 \cdot \frac{4}{\sqrt{576}}\right] \cong [11.67, 12.33]$$

Come cambia la stima se gli stessi dati $\bar{x}, s$ sono ottenuti a partire da un campione composto da 1000 pazienti?

Soluzione: $$\left[12 - 1.96 \cdot \frac{4}{\sqrt{1000}}, 12 + 1.96 \cdot \frac{4}{\sqrt{1000}}\right] \cong [11.75, 12.25]$$

Esercizio 5. Nella seguente tabella sono riportati, raggruppati in classi, i dati relativi al diametro x delle fibre del pelo di un campione di 50 pecore affette da dermatofilosi.

diametro (10^{-1} mm)	f_i
17.75 – 19.75	4
19.75 – 21.75	10
21.75 – 23.75	20
23.75 – 25.75	11
25.75 – 27.75	4
27.75 – 29.75	1
	50

(a) Calcolare media e deviazione standard campionarie, utilizzando la trasformazione $y = \frac{1}{2}(x - 18.75)$.

(b) Rappresentare graficamente la distribuzione delle frequenze.

(c) Rappresentare graficamente la distribuzione delle frequenze cumulate, indicando sull'asse x la posizione della mediana.

(d) Costruire l'intervallo di confidenza al 95% del diametro medio μ nella popolazione.

Esercizi

La trasformazione suggerita $y = \frac{1}{2}(x - 18.75)$ semplifica il calcolo della media e della deviazione standard campionarie:

diametro (10^{-1} mm)	x	f	y	$f \cdot y$	y^2	$f \cdot y^2$
17.75 – 19.75	18.75	4	0	0	0	0
19.75 – 21.75	20.75	10	1	10	1	10
21.75 – 23.75	22.75	20	2	40	4	80
23.75 – 25.75	24.75	11	3	33	9	99
25.75 – 27.75	26.75	4	4	16	16	64
27.75 – 29.75	28.75	1	5	5	25	25
		50		104		278

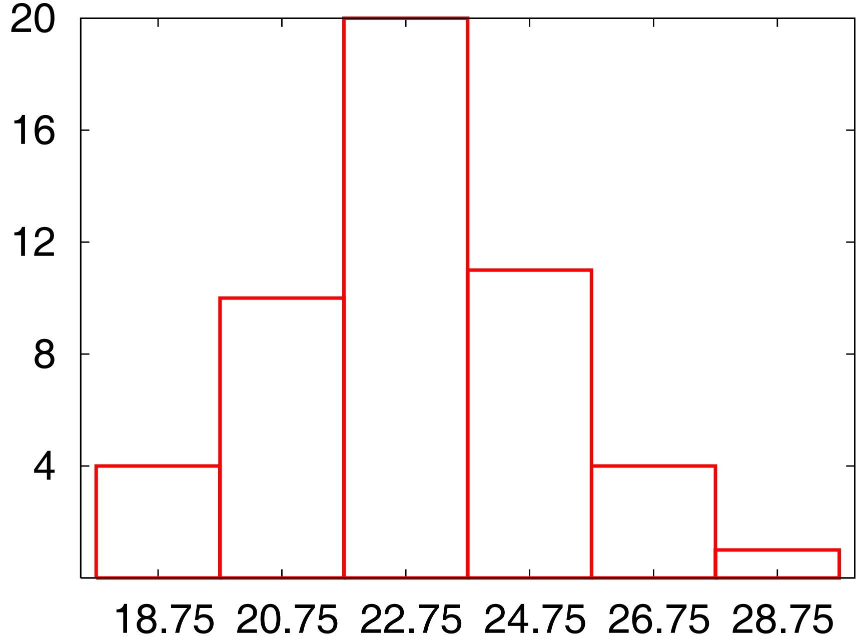
Quindi,

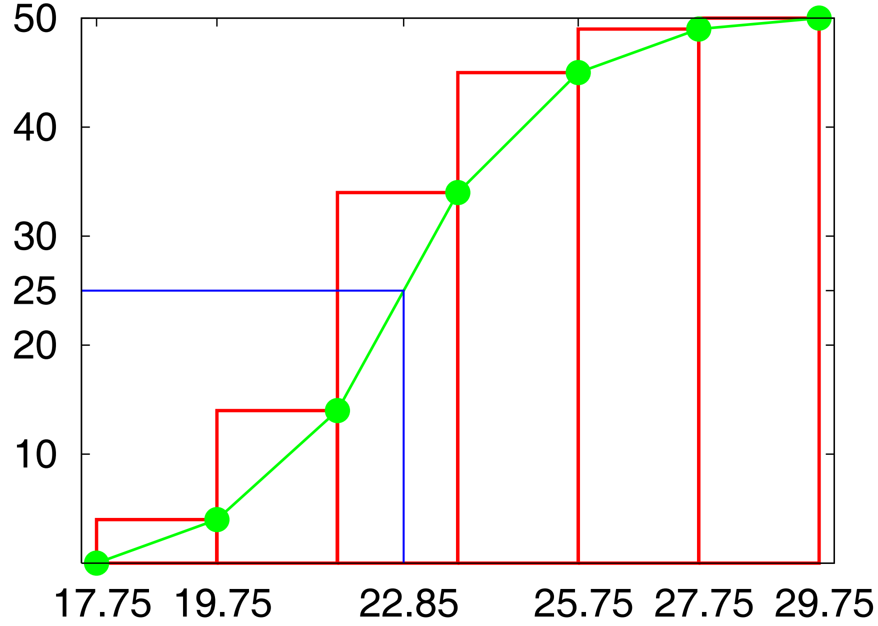
$$\bar{y} = \frac{104}{50} = 2.08 \qquad s_y^2 = \frac{1}{49}\left[278 - 50 \cdot 2.08^2\right] = 1.2588 \qquad s_y = \sqrt{s_y^2} = 1.12$$

Tornando a x mediante la trasformazione inversa $x = 2y + 18.75$,

$$\bar{x} = 2 \cdot (2.08) + 18.75 = 22.91\,(\cdot 10^{-1}\,\text{mm}) \qquad s_x = 2 \cdot (1.12) = 2.24\,(\cdot 10^{-1}\,\text{mm})$$

Esercizi


20
16
12
8
4
18.75 20.75 22.75 24.75 26.75 28.75


50
40
30
25
20
10
17.75 19.75 22.85 25.75 27.75 29.75

Esercizi

Intervallo di confidenza al 95%

Utilizzando la tavola si trova

$$\bar{x} \pm 1.96 \cdot \frac{s_x}{\sqrt{n}} = 22.91 \pm 1.96 \cdot \frac{2.24}{\sqrt{50}} \simeq 22.91 \pm 0.621\,(10^{-1}\,\text{mm})$$

L'intervallo di confidenza al 95% del diametro medio μ (espresso in 10^{-1} mm) è

$$22.91 + 0.621 \leq \mu \leq 22.91 - 0.621$$