come esprimere i dati

arrotondamento :

- **Raccogliere i dati con il maggior numero di cifre significative ed arrotondare eventualmente solo al momento dei calcoli (min. 3);**
- **nella grande maggioranza delle ricerche biologiche l'errore che si commette nel calcolo dei parametri statistici diviene sufficientemente piccolo quando la differenza fra i valori estremi non è inferire a 20-40 unità.**

PESO ALLA NASCITA DEI BOVINI

n	PESO	SESSO
7	38.00	F
8	38.00	F
1	40.00	F
2	40.00	F
5	40.00	F
10	42.00	F
3	47.00	F
9	47.00	M
4	50.00	M
6	50.00	M
CAMPO DI VARIAZIONE	**50.00**	MAX.
	38.00	min.

1- Differenza fra valori estremi:

50 - 38 = 12 unità.

- **Il peso dei vitelli alla nascita avrebbe dovuto essere rilevato con una bilancia più precisa perché la precisione è insufficinete unità<20 !**
- **non sono sufficienti i chili ma sono necessari almeno anche gli etti!**

2- Differenza fra valori estremi:

CAMPO DI VARIAZIONE	**50.20**	MAX.
	37.60	min.

In questo secondo caso 50,2 - 37,6 = 126 unità > 20 la precisione sarebbe stata più che sufficiente.

codifica :

Si può aggiungere o sottrarre da tutti i dati una costante:

- **la media risulta aumentata o diminuita del valore della costante aggiunta o sottratta**
- **i parametri di dispersione**
  - s^2 Varianza **$sx^2/(n-1)$**
  - s o ds Deviazione standard **$\sqrt{s^2}$**
  - c.v. Coefficiente di variazione **s/X**
  - sm o es Errore standard della media **$s/\sqrt{n}$**

non cambiano!

Ricordando che abbiamo definito $sx^2 = S(\overline{X}-X)^2$

Se si aggunge una costante a entrambe le "x" la differenza non cambia

codifica :

Si può moltiplicare o dividere tutti i dati per una costante:

- **la media risulta moltiplicata o divisa per il valore della costante;**
- **il parametro di dispersione "quadrato"**
  - **s^2** Varianza **$s(x-x)^2/(n-1)$ risulta moltiplicato o diviso per il quadrato della costante;**

i parametri di dispersione "riportati lineari"

- **s** Deviazione standard **$\sqrt{s^2}$** e
- **s.m.** Errore standard della media **$s/\sqrt{n}$** **risultano moltiplicati o divisi per la costante;**

il parametro di dispersione

- **c.v.** Coefficiente di variazione **s/X** **non cambia.**

codifiche particolari e metodi di calcolo :

Con le calcolatrici è <u>SEMPRE</u> opportuno:

- **trasformare i numeri decimali in numeri interi**
- **eliminare da tutti i dati le cifre ridondanti.**
- **La somma dei quadrati degli scarti si DEVE ottenere sottraendo alla somma dei quadrati dei dati il "termine di correzione" cioè il quadrato della somma dei dati diviso il numero dei dati della serie.**

$$\sum\left(\bar{x}-x\right)^2=\sum\left(x^2-c\right)$$

dove $C=\dfrac{\sum x^2}{n}$

In parole povere…………..

La:
somma dei quadrati degli scarti
= (è uguale alla)
somma dei quadrati dei dati – (meno la)
somma dei dati "poi" elevata al quadrato e quindi divisa per il numero di dati.

n	valori
1	1
2	2
3	3
media	**2**

metodo "normale" per il calcolo della somma dei quadrati degli scarti $(1-2)^2 + (2-2)^2 + (2-3)^2 = 1+0+1 = \mathbf{2}$

metodo del "termine di correzione"

$1^2 + 2^2 + 3^2 = 1 + 4 + 9 =$ 14-

$\frac{(1+2+3)^2}{3} = \frac{6^2}{3} = \frac{36}{3} =$ 12=

2

Calcolare gli scarti dalla media, elevarli al quadrato, quindi sommarli **è la stessa cosa che** elevare ciascun valore al quadrato, fare la somma dei quadrati, sommarli e quindi sottrarre al valore ottenuto la somma dei dati elevata al quadrato e divisa per il numero dei dati stessi!

Ciò è indispensabile con I dati reali nei quali gli scarti dalla media possono avere molti decimali

Esempio: calcola la deviazione standard della seguente serie di numeri

DATI
3.592
3.447
3.570
3.489
3.202
3.536

	A	B	C	D	E	F	G
1		DATI	ARROTONDATI	CODIFICHE		CODIFICATI	quadrati
2		3,592	=ARROTONDA(B2;2)	-3	*100	=+(C2+D2)*100	=+F2^2
3		3,447	=ARROTONDA(B3;2)	-3	*100	=+(C3+D3)*100	=+F3^2
4		3,57	=ARROTONDA(B4;2)	-3	*100	=+(C4+D4)*100	=+F4^2
5		3,489	=ARROTONDA(B5;2)	-3	*100	=+(C5+D5)*100	=+F5^2
6		3,202	=ARROTONDA(B6;2)	-3	*100	=+(C6+D6)*100	=+F6^2
7		3,536	=ARROTONDA(B7;2)	-3	*100	=+(C7+D7)*100	=+F7^2
8	MAX	=MAX(B2:B7)	=MAX(C2:C7)				
9	min	=MIN(B2:B7)	=MIN(C2:C7)				
10	intervallo	=+(B8-B9)*1000	=+(C8-C9)*100				
11	n	=CONTA.VALORI(B2:					
12	somma		=SOMMA(C2:C7)	=+F12^2		=SOMMA(F2:F7)	=SOMMA(G2:G7)
13	media	=MEDIA(B2:B7)	=MEDIA(C2:C7)	=+D12/B11		=MEDIA(F2:F7)	
14			Somma quadrati scarti =	=+G12-D13		100/	
15			g.l.=	=+B11-1		3,00+	
16			Varianza codif.=	=+D14/D15		=+F13/100+3	
17	d.s. codificata			=RADQ(D16)		=DEV.ST(F2:F7)	
18				/100		/100	
19	d.s.	=DEV.ST(B2:B7)	=DEV.ST(C2:C7)	=+D17/100		=+F17/100	
20							

	DATI	ARROTONDATI		CODIFICHE	CODIFICATI		quadrati
	3.592	3.59	-3.00	*100	59		3481
	3.447	3.45	-3.00	*100	45		2025
	3.570	3.57	-3.00	*100	57		3249
	3.489	3.49	-3.00	*100	49		2401
	3.202	3.20	-3.00	*100	20		400
	3.536	3.54	-3.00	*100	54		2916
MAX	3.592	3.59					
min	3.202	3.20					
intervallo	390	39					
			quadrato della somma				
n	6						
somma		20.84	80656	← al quadrato	**284**		14472.00
media	3.4727	3.4733	13442.667	= quadrato della somma diviso n=6	47.33	**=media codificata**	
			1029.33	= somma dei quadrati degli scarti	100/		
		g.l.=	5		3,00+		
		Varianza =	205.86667		3.4733	**=media riportata a valori originari**	
			14.3481		14.3481	**= d.s. codificata**	
			/100		/100		
d.s.	0.142766	0.143481	0.143481		0.143480545	**= d.s. decodificata**	

Presentazione risultati e arrotondamento finale

Il numero di decimali della media dipende dalla dispersione dei dati cioè dalla s, d.s., e.s.

REGOLA =

- **La media deve essere arrotondata al decimo del suo e.s.** (errore standard)**.**
- **La d.s. o e.s.** (deviazione standard, errore standard) **devono riportare una cifra decimale in più della media**

Troppi decimali = falsa impressione di precisione che in realtà non c'è **(errore lieve)**;

Pochi decimali = perdita informazione e impossibilità di utilizzo dei risultati da pare di terzi **(errore grave).**

Esempio

	Dati bruti	Dati arrotondati correttamente
Media =	3.4727	**3.473**
d.s. =	0.142766	**0.1428**
e.s. =	0.0582842	0.0583

1- Prima cifra "non zero" ai centesimi

3- elemento di dispersione con cifre fino ai decimillasimi = una posizione decimale in più della media "millesimi+1"

2- Media con cifre fino ai millesimi "centesimi+1"

Esempi

Riporta correttamente gli elementi che descrivono il campione

	originali	originali	originali
n =	3	300	3
Media =	3.4726666667	3.4726666667	3.4726666667
d.s. =	0.176647599	0.176647599	1.7664759902
calcolo l'e.s.	prima cifra al decimo	prima cifra al centesimo	prima cifra alla unità
e.s. =	0.1019875388	0.0101987539	1.0198753885

	originali	correttamente espressi	originali	correttamente espressi	originali	correttamente espressi
n	3	3	300	300	3	3
media	3,472666667	3,47	3,472666667	3,473	3,472666667	3,5
d.s.	0,176647599	0,177	0,176647599	0,1766	1,76647599	1,77
	prima cifra al decimo		prima cifra al centesimo		prima cifra ala unità	
e.s.	0,101987539		0,010198754		1,019875388	

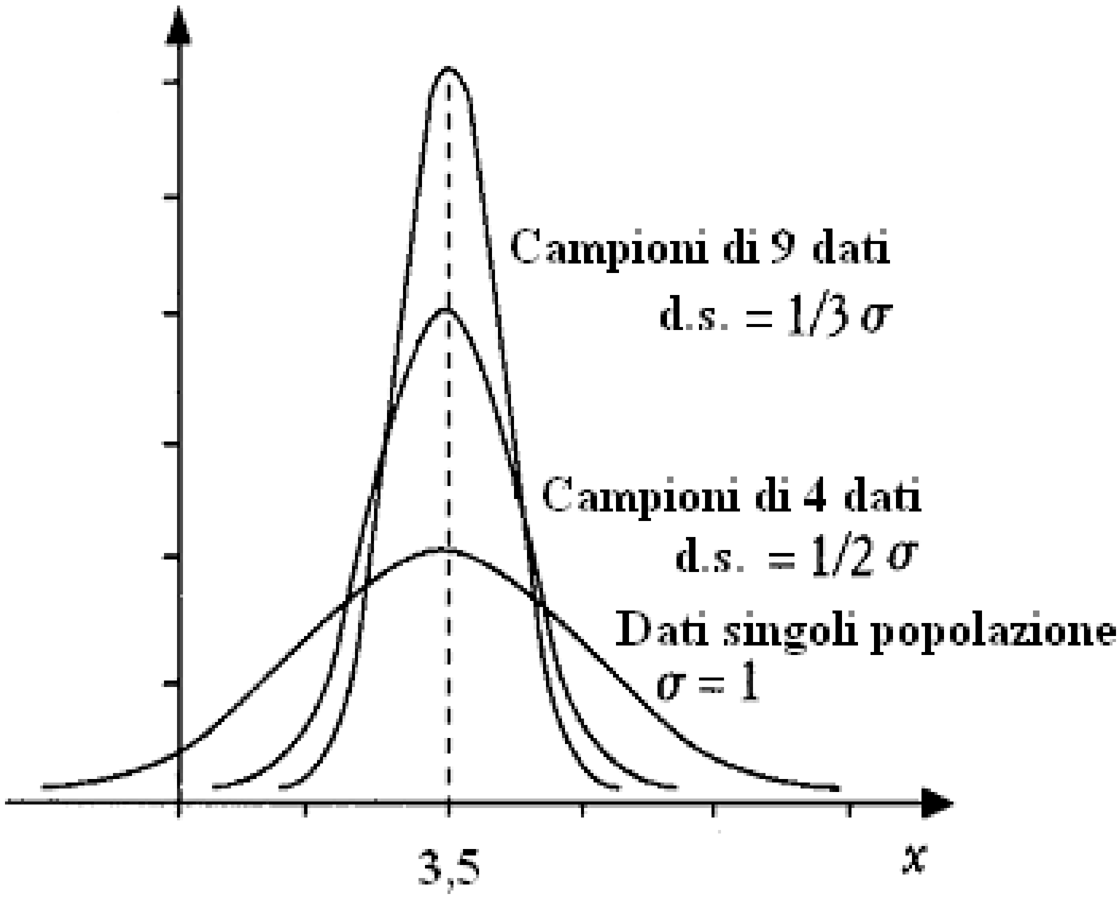
Le medie calcolate a partire da un campione oscillano molto meno attorno alla media di popolazione di quanto non fanno le singole misure

Quanto maggiore è la dimensione del campione (numero di misure di un campione) tanto minore è la dispersione.

Abbiamo infatti più fiducia in una media di un campione che in una misura isolata!

Se una **popolazione** è **distribuita normalmente** con una media di popolazione **μ** e una deviazione standard di popolazione **σ**, le **medie** di un numero infinito **di campioni -** <u>ciascuno composto da n individui</u> estratti a caso dalla popolazione - si distribuiscono secondo una distribuzione normale le cui **medie sono la media della popolazione** e le cui **deviazioni standard saranno l'errore standard** cioè **d.s. = $\sigma/\sqrt{n}$**

Ricorda la deviazione standard è uguale alla radice quadrata della varianza, quindi un numero (la media) che deriva **"n"** osservazioni sarà pari alla varianza diviso "**n**" da cui, estratta la radice quadrata, la deviazione standard della popolazione diviso la radice di n = individui presenti nel campione.



POPOLAZIONE		CAMPIONI		
n	infinito	4	9	16
media	3,5			
deviazione standard	0,15	0,075	0,05	0,0375
rapporto con deviazione standard		1/2	1/3	1/4
		0,075	0,05	0,0375

L'errore standard varia in ragione della radice quadrata della numerosità del campione, cioè di $1/\sqrt{n}$

Le considerazioni che precedono sono **valide** solo **nel caso** che **si conosca** la **media** μ e la **deviazione standard** δ **della popolazione**.

attenzione!

Non è lecito fare i calcoli a partire dalla media e dalla deviazione standard di un campione!

Queste sono infatti solo una stima dei valori originali e per questo possono essere sbagliate; o meglio, non sono esattamente gli stessi valori dell'originale ma assumono valori che si discostano da quello vero, presentano cioè degli errori di stima in più o in meno che devono essere considerati.

da **Popolazione** a **Campione**

A partire dalla popolazione di origine è **impossibile determinare esattamente** un campione estratto dalla popolazione, cioè conoscerne la $\overline{x}$ (media aritmetica) e la d.s. (deviazione standard)

A partire dalla popolazione di origine è **possibile stimare i valori più probabili** di un campione estratto dalla popolazione, cioè conoscerne la $\overline{x}$ (media aritmetica) e la d.s. (deviazione standard)

Ho la seguente popolazione:

POPOLAZIONE	
n	infinito
media	3.5
deviazione standard	0.15

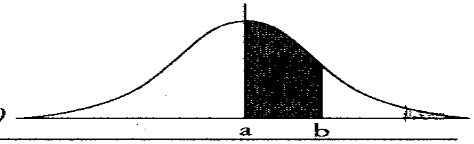
come può variare la media di un campione di 6 misure estratte da tale popolazone?

Es. calcola entro quali valori ricade la media del campione nel 95,44% dei casi

Da Tavole probabilità

95,44% / 2 = 47,72% = 2 z

TAV. B. *Aree della distribuzione normale standard tra* $a = 0$ *e* $b > 0$



z	0	1	2	3	4	5	6	7	8	9
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0754
0.2	.0793	8.032	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2258	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2518	.2549
0.7	.2580	.2612	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2996	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4351	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998	.4998
3.6	.4998	.4998	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.7	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.8	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999	.4999
3.9	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000	.5000

Fonte: Cristante, Lis, Sambin [1980].

La media del campione si distribuisce secondo una distribuzione normale le cui **media** è **3,5** e la cui **deviazione standard è l'errore standard**

cioè **= σ/√n = 0,15/√6** **= 6 misure**

per cui 2 deviazioni standard =

	POPOLAZIONE	CAMPIONE	
n	infinito	6	
media	3,5	3,5	
deviazione standard	0,15	0,0612372436	per 2 =
	2σ	0,1224744871	da cui +o- la media
max		3,6224744871	
min		3,3775255129	

nel 95,44% dei casi la media del campione sarà compresa fra

Nella maggior parte dei casi però noi non conosciamo in anticipo né la media né la deviazione standard della popolazione!
È proprio per avere questa informazione che facciamo le misurazioni di un campione!

cioè

Nella maggior parte dei casi il nostro **non è un caso di probabilità** (estrazione) ma **è un caso di inferenza** (si congetturano le caratteristiche di una popolazione animale a partire dai parametri determinati su un campione).

vedi lezione

La distribuzione del "t" di Student