

Le basi dell'inferenza statistica

Massimo Borelli

18 luglio 2024





inferenza: il concetto generale



che cosa è “S.E. mean”



un “giochetto”

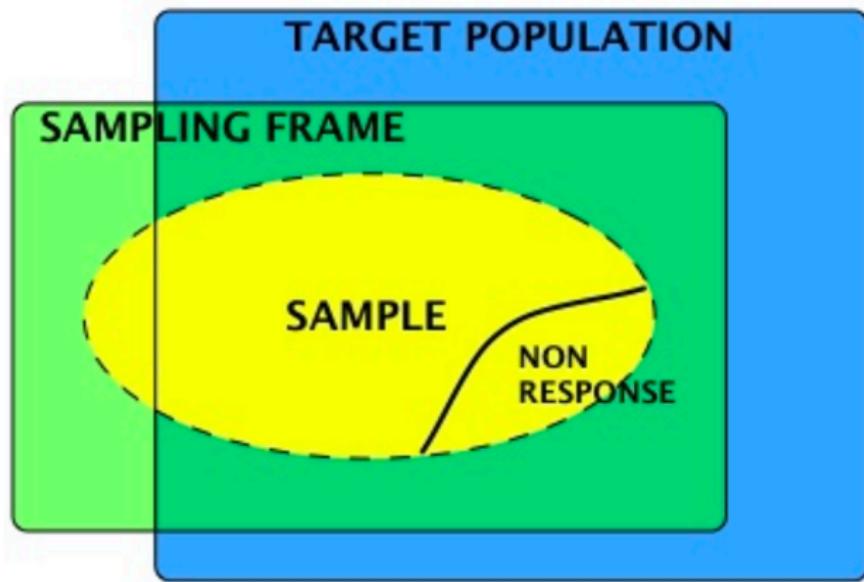


la grande scoperta



non sbagliare mai più

inferenza: il concetto generale



http://www.scholarpedia.org/article/Sampling_bias

inferenza: il concetto generale



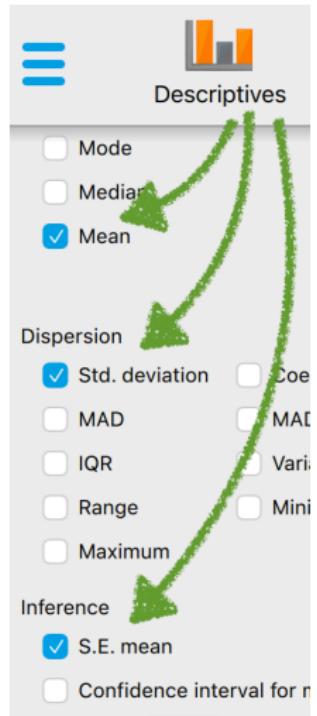
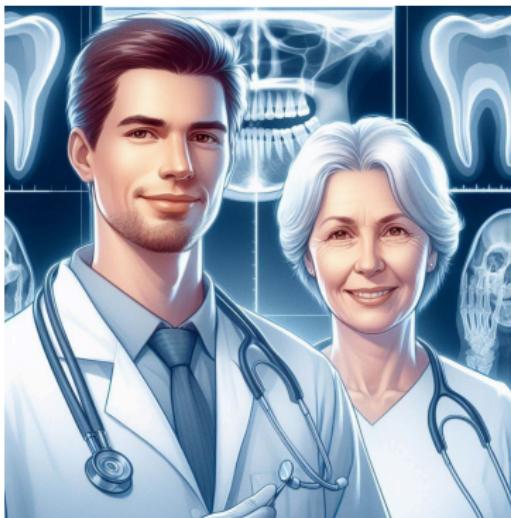
.. persino una vecchia versione di JASP

Dispersion

<input checked="" type="checkbox"/> S.E.mean	<input checked="" type="checkbox"/> Std.deviation
<input type="checkbox"/> Coefficient of Variation	<input type="checkbox"/> MAD
<input type="checkbox"/> MAD Robust	<input type="checkbox"/> IQR
<input type="checkbox"/> Variance	<input type="checkbox"/> Range
<input type="checkbox"/> Minimum	<input type="checkbox"/> Maximum

perché?

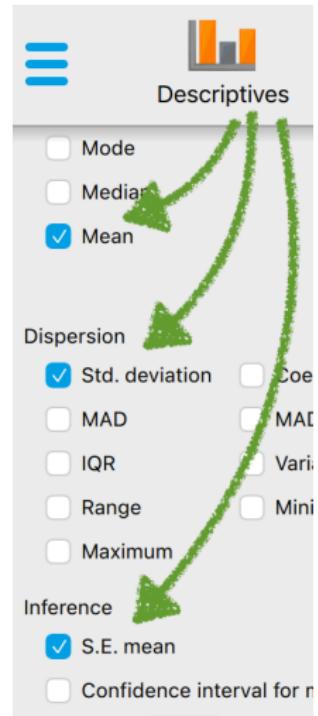
il dataset affidabilita.ods



perché?

il dataset affidabilita.ods

	etapaziente	
	dottore	professoressa
Valid	100	1000
Mean	43.220	43.663
Std. Error of Mean	1.440	0.445
Std. Deviation	14.397	14.077



un gioco: non confondere variabilità con affidabilità



47 9 37 50 45



63 48 46 14 24

un gioco: non confondere variabilità con affidabilità



media 37.6

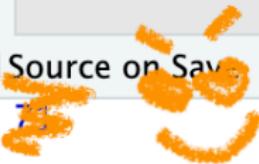


media = 39

un gioco: non confondere variabilità con affidabilità

media							media						
68	42	4	25	27	33,2		32	46	59	40	51	45,6	
27	45	39	47	64	44,4		52	67	14	59	21	42,6	
60	25	52	43	34	42,8		36	56	30	47	67	47,2	
39	10	59	69	42	43,8		14	7	24	21	9	15,0	
61	31	2	66	24	36,8		23	68	13	4	69	35,4	
21	57	41	6	9	26,8		61	21	44	53	71	50,0	
54	43	9	49	68	44,6		39	16	64	27	68	42,8	
7	55	36	49	17	32,8		63	43	21	10	61	39,6	
25	37	20	49	40	34,2		27	67	51	8	54	41,4	
63	8	69	3	29	34,4		51	15	26	34	69	39,0	
media delle medie 37,4							media delle medie 39,9						

un gioco: non confondere variabilità con affidabilità



The screenshot shows an RStudio environment with the following details:

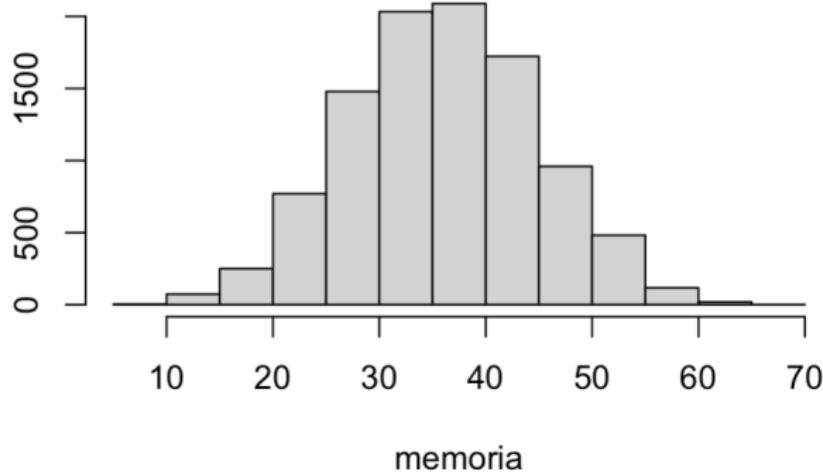
- Title Bar:** Shows the file name "02winference.R".
- Toolbar:** Includes standard icons for back, forward, save, and search.
- Text Editor:** Displays the R code:

```
1 secreto = 7
2
3 memoria = numeric(10000)
4
5 for (i in 1:10000)
6   memoria[i] = mean(sample(1:secreto, 5))
7
8 mean(memoria)
9
```

un gioco: non confondere variabilità con affidabilità

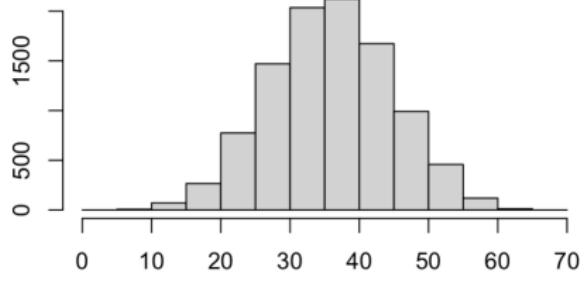
```
02winference.R x
Source on Save
1 secreto = 36
2
3 memoria = numeric(10000)
4
5 for (i in 1:10000)
6   memoria[i] = mean(sample(1:secreto, 5))
7
8 mean(memoria)
a
```

5 palline estratte, media ~ 36



un gioco: non confondere variabilità con affidabilità

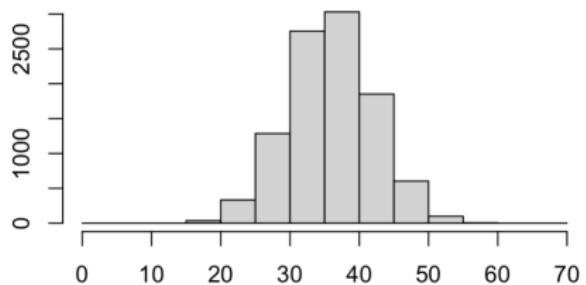
5 palline estratte, media ~ 36



47 9 37 50 45



10 palline estratte, media ~ 36



40 29 7 52 58 30 11 56 9 8



Cosa possiamo osservare?

- ➊ che le medie (36) coincidono con la media vera (36)
- ➋ che la distribuzione della media è gaussiana
- ➌ e che .. *la campana si stringe* ..?

un gioco: non confondere variabilità con affidabilità

TABELLONE									
1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71									

Console Terminal

R 4.4.0 · ~/Docu

```
> mean(1:71)
[1] 36
> sd(1:71)
[1] 20.6
```

esperimento	media	sd
(bambini) 5 palline	36	9.0
(nonni) 10 palline	36	6.5

Come si 'collegano' i numeri 6.5, 9.0 e 20.6?

GRANDE SCOPERTA SCIENTIFICA

Le leggi dei gradi numeri, il teorema centrale del limite, ...

esperimento	media	sd	\sqrt{n}	$sd \cdot \sqrt{n}$
(bambini) 5 palline	36	9.0	2.24	20.1
(nonni) 10 palline	36	6.5	3.16	20.8

- ➊ che le medie (36) coincidono con la media vera (36)
- ➋ che la distribuzione della media è gaussiana
- ➌ e che le deviazioni standard della media campionaria sono in relazione con la deviazione standard della popolazione

l'errore standard della media

$$sem = \frac{\sigma}{\sqrt{n}}$$

Console

Terminal

R 4.4.0 · ~/Docu
> mean(1:71)
[1] 36
> sd(1:71)
[1] 20.6

esperimento	media	sem	\sqrt{n}
(bambini) 5 palline	36	9.0	2.24
(nonni) 10 palline	36	6.5	3.16

ecco spiegato il perché!

	etapaziente	
	dottore	professoressa
Valid	100	1000
Mean	43.220	43.663
Std. Error of Mean	1.440	0.445
Std. Deviation	14.397	14.077



$$1.44 = \frac{14.397}{\sqrt{100}}$$

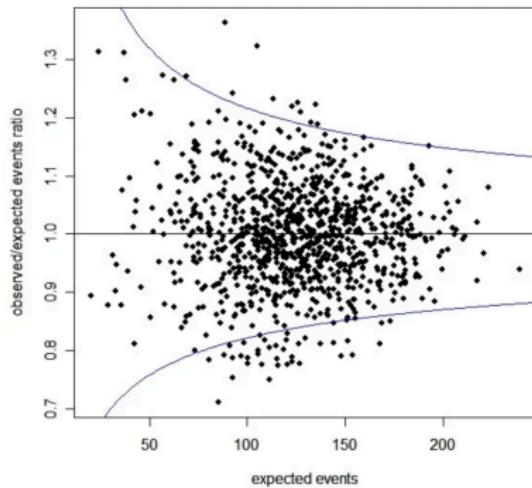
$$0.445 = \frac{14.077}{\sqrt{1000}}$$

Purtroppo, in passato, c'è stata confusione ..

Yu-Kang Tu and Mark Gilthorpe.

The most dangerous hospital or the most dangerous equation?

<https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-7-185>



Purtroppo, in passato, c'è stata confusione ..

Howard Weiner.

The Most Dangerous Equation

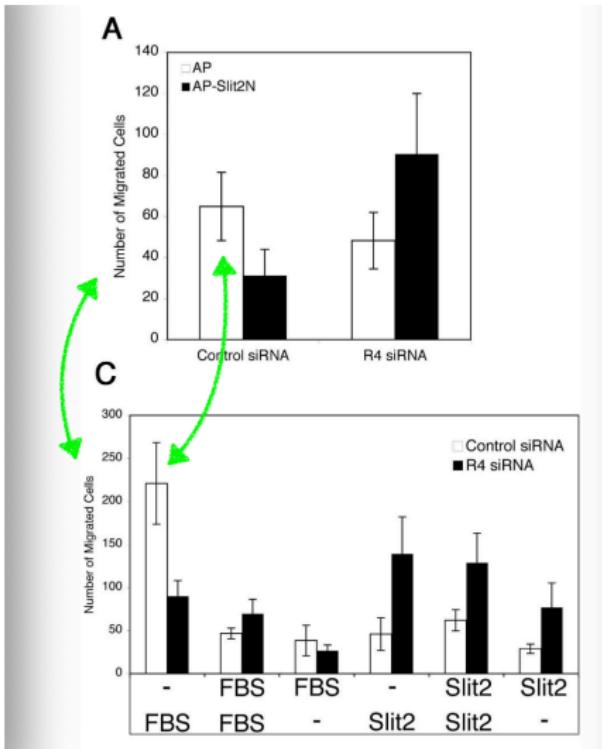
https://nexthumanproject.com/references/Y_Variant_Of_XYChromosome.pdf



Purtroppo, in passato, c'è stata confusione ..

BMC Cell Biology

Research article
Silencing of directional migration in *roundabout4* knockdown endothelial cells
Sukhbir Kaur^{†,2},



Purtroppo, in passato, c'è stata confusione ..

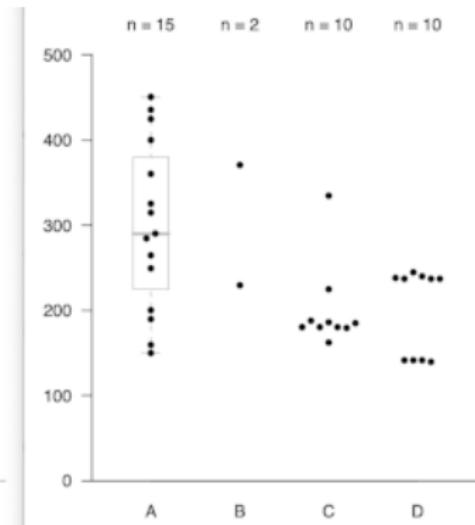
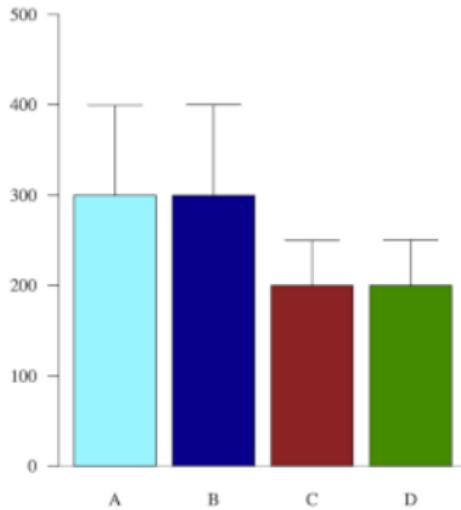
transfected endothelial cells. **C** shows migration assay for control *iacc2* and *robo4* siRNA transfected cells to Serum or AP-Slit2N in either upper (U), lower (L) or both chambers as indicated. Error bars in **A** ($n = 3$), and **B** ($n = 3$) represent SD while in **C** represent SEM ($n = 4$). **D** shows pulldown analysis of Cdc42-GTP levels in AP and AP-Slit2N (25 ng/ml) treated endothelial cell lysates for 0 and 15 minutes respectively. β -tubulin

Lettura consigliata

Tatsuki Koyama.

Beware of Dynamite

<https://biostat.app.vumc.org/wiki/pub/Main/TatsukiRcode/Poster3.pdf>



Take home message

- la **deviazione standard** misura la **variabilità** dei dati del *nostro campione*
 - diciamola a tutti, è utile!
- l'**errore standard** misura l'**affidabilità** della *nostra congettura* che la media m del *nostro campione* catturi la media μ (ignota) della popolazione
 - non diciamola a nessuno, confonde!
 - + eccezione: 'duplicato', 'triplicato', ..
 - occhio alla dinamite!

