

L'analisi della varianza

Rappresentazione grafica e calcoli col foglio
elettronico

Giovanni Battista Flebus

Per ogni riga (soggetto)...

- Si considera il valore osservato
- Si calcola la differenza dalla media
- Si calcola il suo quadrato

Alla fine...

- Si esegue la somma totale dei quadrati
- Si divide per i gradi di libertà
- Si ottiene la stima della varianza della popolazione

stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato
s1	A	2	4	5,67	-3,67	13,44
s2	A	4	4	5,67	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11
s5	B	4	6	5,67	-1,67	2,78
s6	B	5	6	5,67	-0,67	0,44
s7	B	7	6	5,67	1,33	1,78
s8	B	8	6	5,67	2,33	5,44
s9	C	5	7	5,67	-0,67	0,44
s10	C	7	7	5,67	1,33	1,78
s11	C	8	7	5,67	2,33	5,44
s12	C	8	7	5,67	2,33	5,44
somma		68	68	68	0,00	42,67
gradi di liberta						11
varianza						3,88

Devianza totale.
 Serve per i controlli

42,67

							nei gruppi		fra	
stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato	scarto dal gruppo	quadrato	scarto dal totale	quadrato
s1	A	2	4	5,67	-3,67	13,44	-2	4	-1,67	2,78
s2	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11	2	4	-1,67	2,78
s5	B	4	6	5,67	-1,67	2,78	-2	4	0,33	0,11
s6	B	5	6	5,67	-0,67	0,44	-1	1	0,33	0,11
s7	B	7	6	5,67	1,33	1,78	1	1	0,33	0,11
s8	B	8	6	5,67	2,33	5,44	2	4	0,33	0,11
s9	C	5	7	5,67	-0,67	0,44	-2	4	1,33	1,78
s10	C	7	7	5,67	1,33	1,78	0	0	1,33	1,78
s11	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
s12	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
somma		68	68	68	0,00	42,67	0	24	0,00	18,67
gradi di liberta						11			9	2
varianza						3,88			2,67	9,33

Devianza
nei gruppi

Devianza
fra i gruppi

							nei gruppi		fra	
stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato	scarto dal gruppo	quadrato	scarto dal totale	quadrato
s1	A	2	4	5,67	-3,67	13,44	-2	4	-1,67	2,78
s2	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11	2	4	-1,67	2,78
s5	B	4	6	5,67	-1,67	2,78	-2	4	0,33	0,11
s6	B	5	6	5,67	-0,67	0,44	-1	1	0,33	0,11
s7	B	7	6	5,67	1,33	1,78	1	1	0,33	0,11
s8	B	8	6	5,67	2,33	5,44	2	4	0,33	0,11
s9	C	5	7	5,67	-0,67	0,44	-2	4	1,33	1,78
s10	C	7	7	5,67	1,33	1,78	0	0	1,33	1,78
s11	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
s12	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
somma		68	68	68	0,00	42,67	0	24	0,00	18,67
gradi di liberta						11	9	2		
varianza						3,88	2,67	9,33		

							nei gruppi		fra	
stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato	scarto dal gruppo	quadrato	scarto dal totale	quadrato
s1	A	2	4	5,67	-3,67	13,44	-2	4	-1,67	2,78
s2	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11	2	4	-1,67	2,78
s5	B	4	6	5,67	-1,67	2,78	-2	4	0,33	0,11
s6	B	5	6	5,67	-0,67	0,44	-1	1	0,33	0,11
s7	B	7	6	5,67	1,33	1,78	1	1	0,33	0,11
s8	B	8	6	5,67	2,33	5,44	2	4	0,33	0,11
s9	C	5	7	5,67	-0,67	0,44	-2	4	1,33	1,78
s10	C	7	7	5,67	1,33	1,78	0	0	1,33	1,78
s11	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
s12	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
somma		68	68	68	0,00	42,67	0	24	0,00	18,67
gradi di liberta						11	9	2		
varianza						3,88	2,67	9,33		

	Somma dei quadrati	df	Media dei quadrati	F	Sig.
Fra gruppi	18,667	2	9,333	3,500	,075
Entro gruppi	24,000	9	2,667		
Totale	42,667	11			

stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato	nei gruppi		fra	
							scarto dal gruppo	quadrato	scarto gruppo dal totale	quadrato
s1	A	2	4	5,67	-3,67	13,44	-2	4	-1,67	2,78
s2	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11	2	4	-1,67	2,78
s5	B	4	6	5,67	-1,67	2,78	-2	4	0,33	0,11
s6	B	5	6	5,67	-0,67	0,44	-1	1	0,33	0,11
s7	B	7	6	5,67	1,33	1,78	1	1	0,33	0,11
s8	B	8	6	5,67	2,33	5,44	2	4	0,33	0,11
s9	C	5	7	5,67	-0,67	0,44	-2	4	1,33	1,78
s10	C	7	7	5,67	1,33	1,78	0	0	1,33	1,78
s11	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
s12	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
somma		68	68	68	0,00	42,67	0	24	0,00	18,67
gradi di liberta						11	9	2		
varianza						3,88	2,67	9,33		

	Somma dei quadrati	df	Media dei quadrati	F	Sig.
Fra gruppi	18,667	2	9,333	3,500	,075
Entro gruppi	24,000	9	2,667		
Totale	42,667	11			

Il rapporto F

- La statistica F è quindi un rapporto fra due varianze, calcolate dividendo la devianza fra i gruppi per la devianza nei gruppi, ognuna divisa per i rispettivi gradi di libertà

$$\frac{SQ_{fra} / (k - 1)}{SQ_{dentro} / (n - k)}$$

F di Fisher- Snedecor

$$F = \frac{X/m}{Y/n}$$

ove X e Y sono due variabili casuali indipendenti con rispettive distribuzioni chi quadrato con m ed n *gradi di libertà*

Due usi per il termine gradi di libertà

- 1 Denominatore per calcolare la varianza della popolazione
- 2 Conteggio dei parametri per le distribuzioni come il chi quadrato e l'F di Fisher-Snedecor

I gradi di libertà

Ad ognuna delle devianze sono associati i gradi di libertà:

- la **devianza totale** ha $n - 1$ gradi di libertà
- la **devianza tra gruppi** ha $k - 1$ gradi di libertà
- la **devianza entro i gruppi** ha $n - p$ gradi di libertà

Dividendo ciascuna devianza *per* i rispettivi gradi di libertà si ottengono le media dei quadrati, cioè le **VARIANZE**:

$$\text{var } fra = \frac{SQF}{p - 1}$$

$$\text{var } err = \frac{SQE}{n_t - p}$$

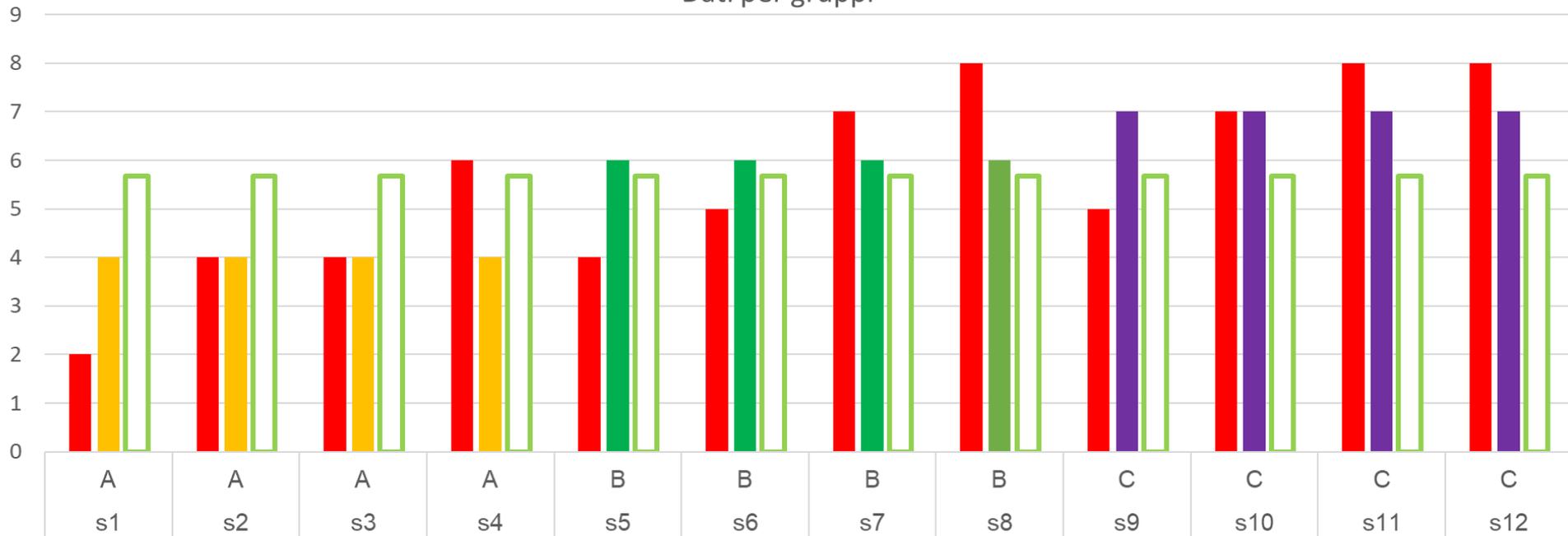
Varianza tra i gruppi

Varianza entro i gruppi

stud	gruppo	N esami	medie gruppo	media totale	scarto dalla media totale	quadrato	nei gruppi		fra	
							scarto dal gruppo	quadrato	scarto gruppo dal totale	quadrato
s1	A	2	4	5,67	-3,67	13,44	-2	4	-1,67	2,78
s2	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s3	A	4	4	5,67	-1,67	2,78	0	0	-1,67	2,78
s4	A	6	4	5,67	0,33	0,11	2	4	-1,67	2,78
s5	B	4	6	5,67	-1,67	2,78	-2	4	0,33	0,11
s6	B	5	6	5,67	-0,67	0,44	-1	1	0,33	0,11
s7	B	7	6	5,67	1,33	1,78	1	1	0,33	0,11
s8	B	8	6	5,67	2,33	5,44	2	4	0,33	0,11
s9	C	5	7	5,67	-0,67	0,44	-2	4	1,33	1,78
s10	C	7	7	5,67	1,33	1,78	0	0	1,33	1,78
s11	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
s12	C	8	7	5,67	2,33	5,44	1	1	1,33	1,78
somma		68	68	68	0,00	42,67	0	24	0,00	18,67
-----gradi di liberta -----					11		9		2	
varianza						3,88		2,67		9,33

Rappresentazione grafica di punteggi, scarti dalla media e devianza

Dati per gruppi

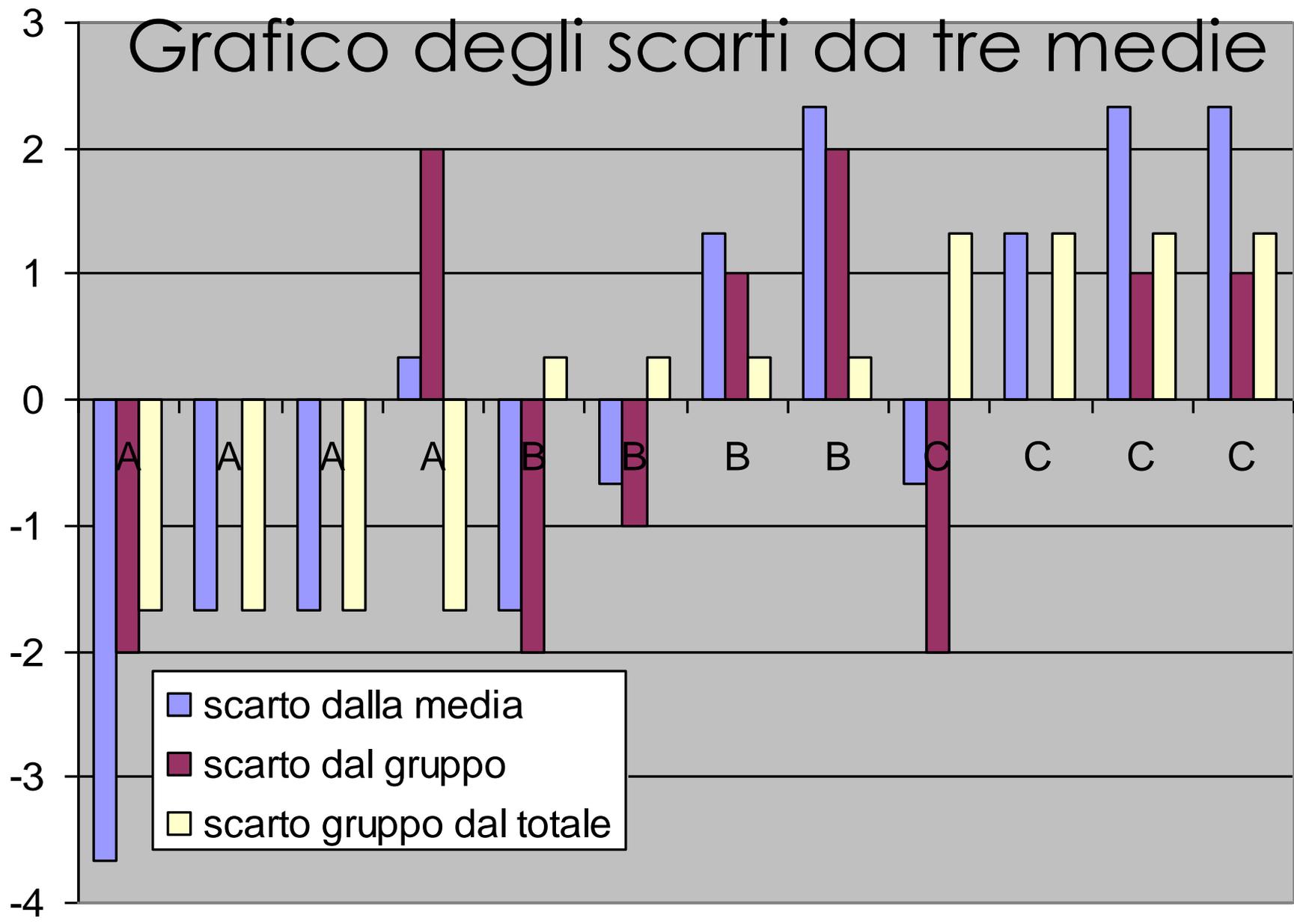


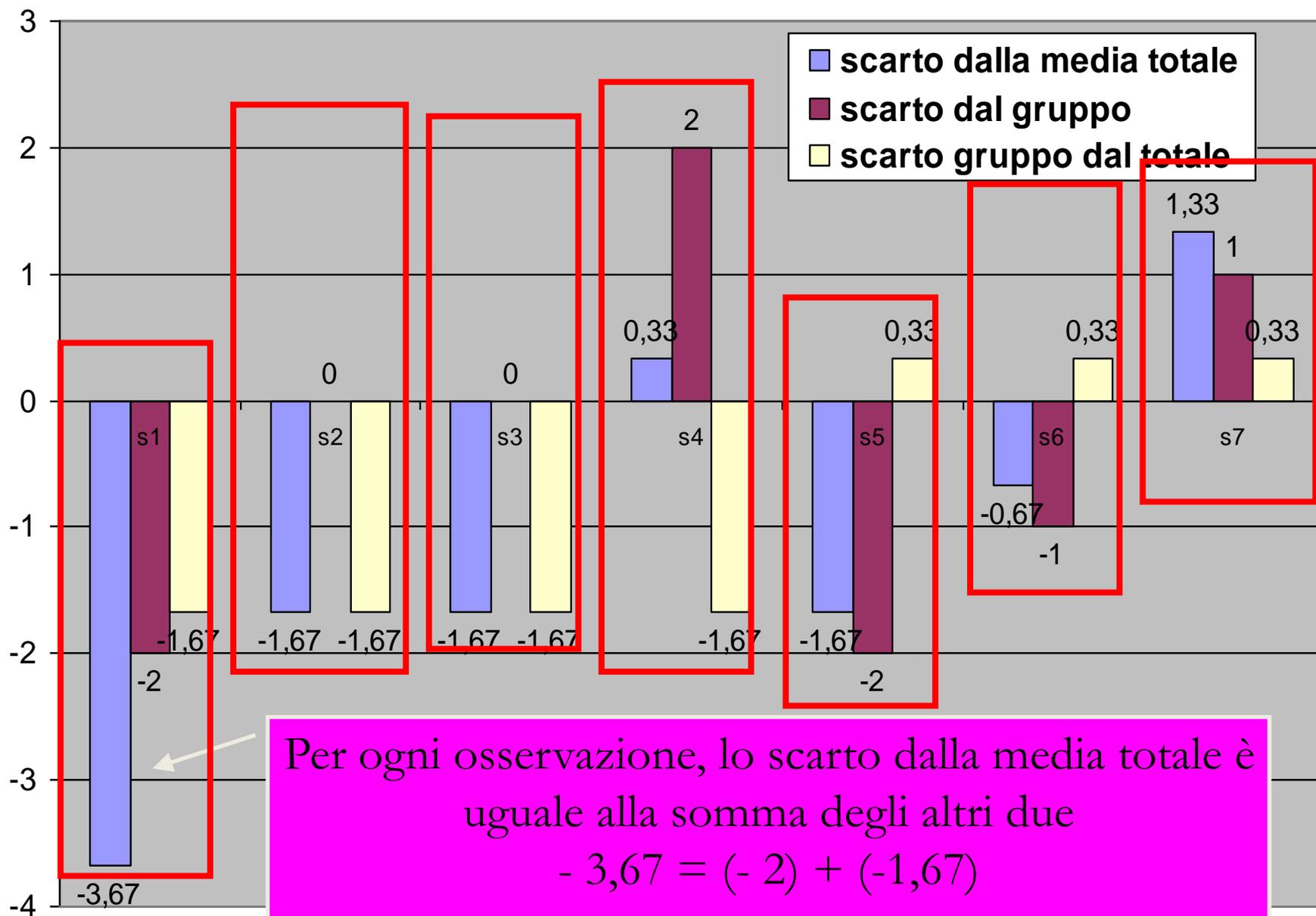
Rosso : osservazione per individuo

Giallo o verde o viola: media del gruppo

Bianco bordato di verde: media totale

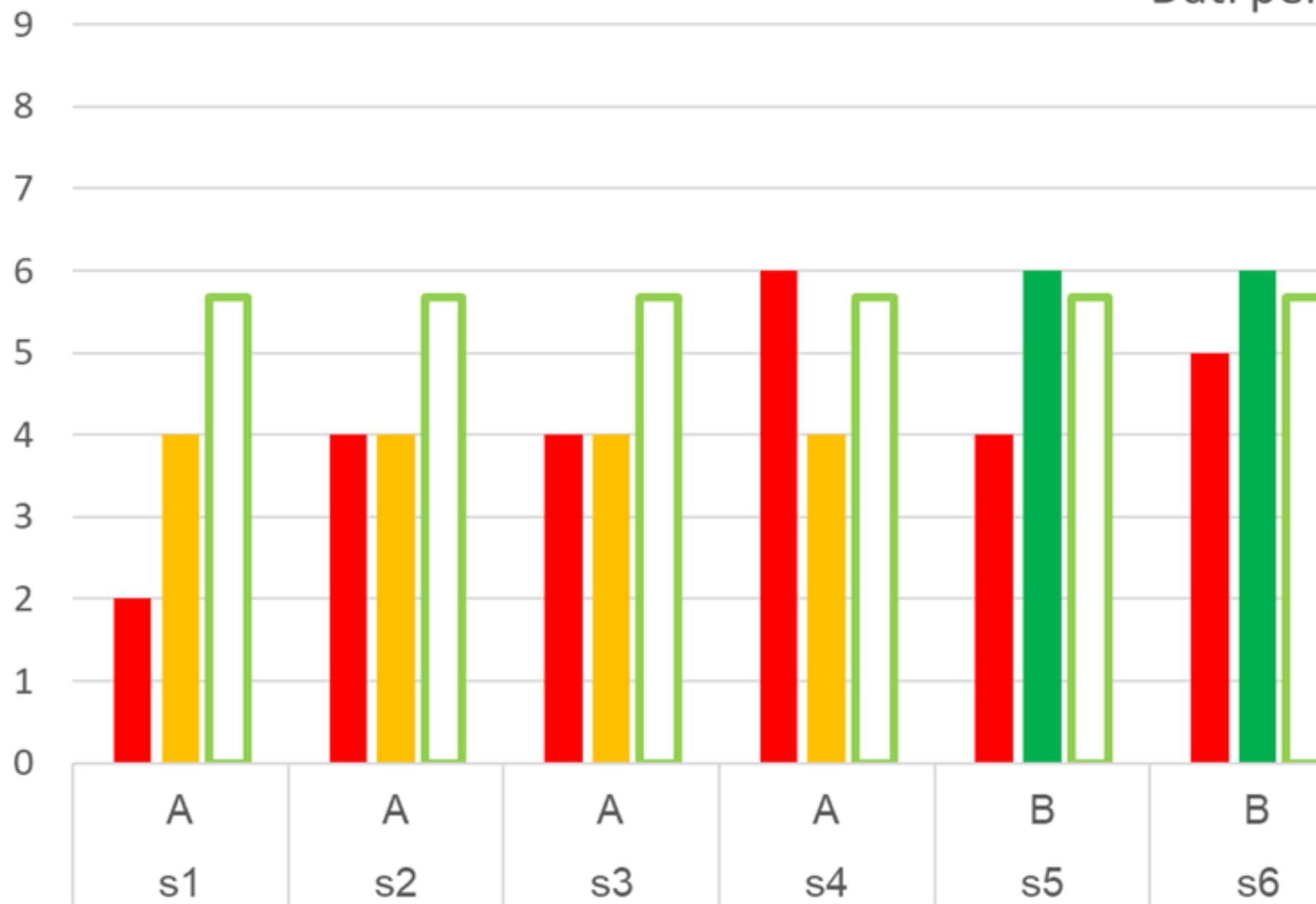
Grafico degli scarti da tre medie





Torniamo al grafico precedente

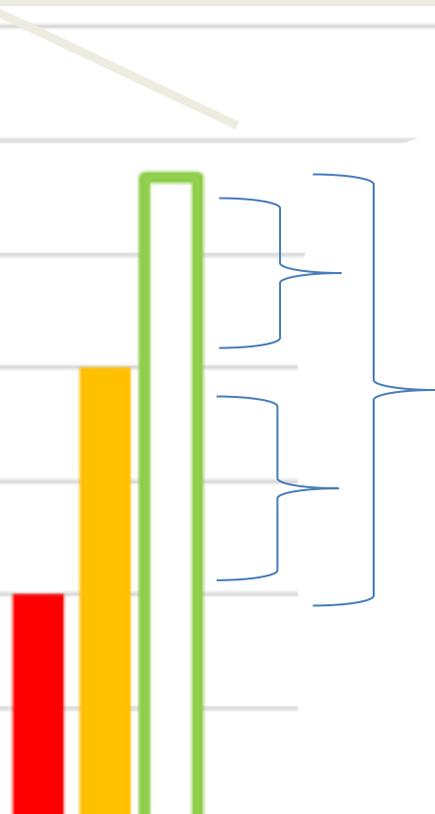
- Esaminiamo il primo studente, che ha un
- Numero di esami pari a 2
- La media del suo gruppo è 4
- La media dell'intero campione è pari a 5,67

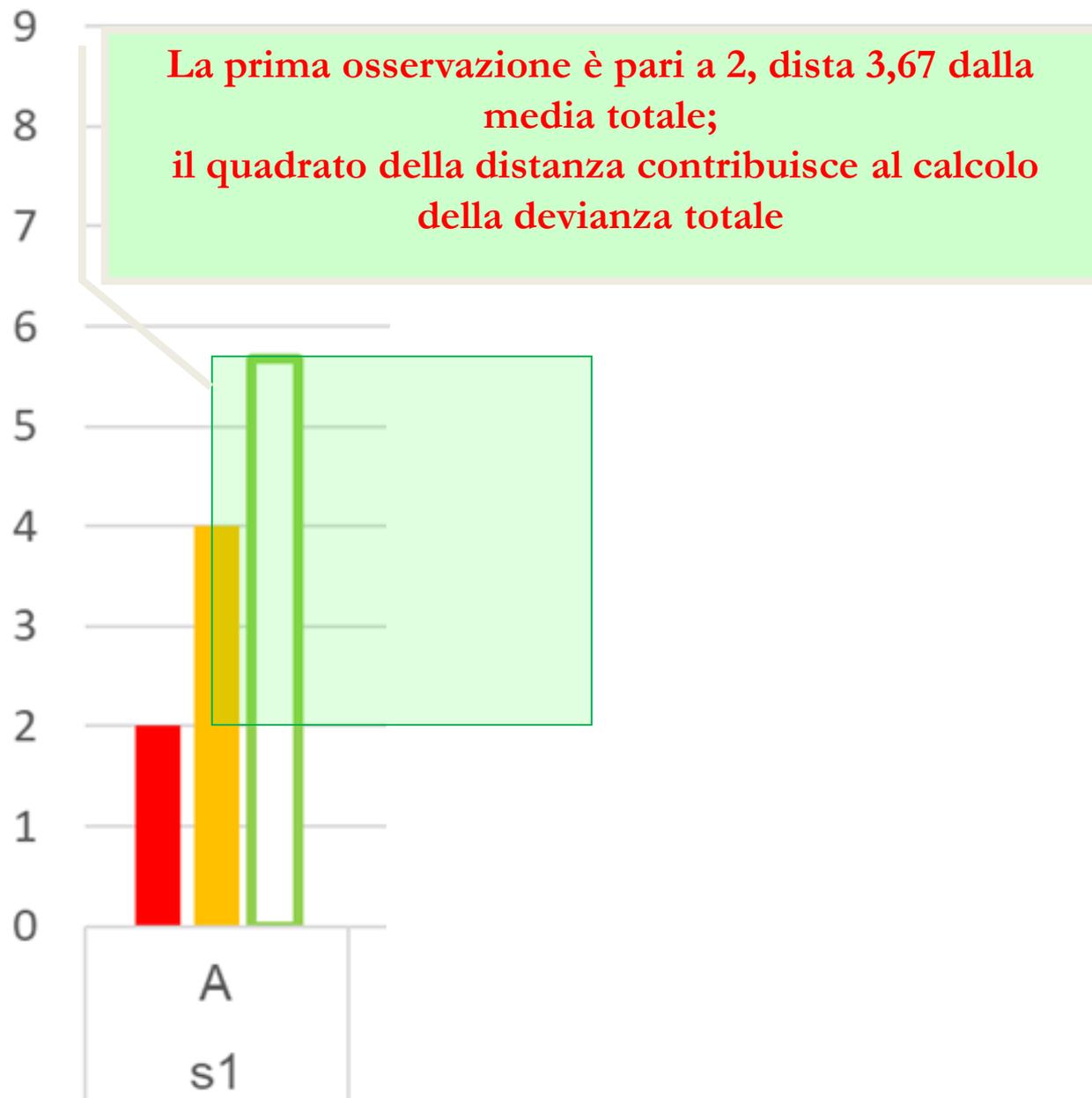


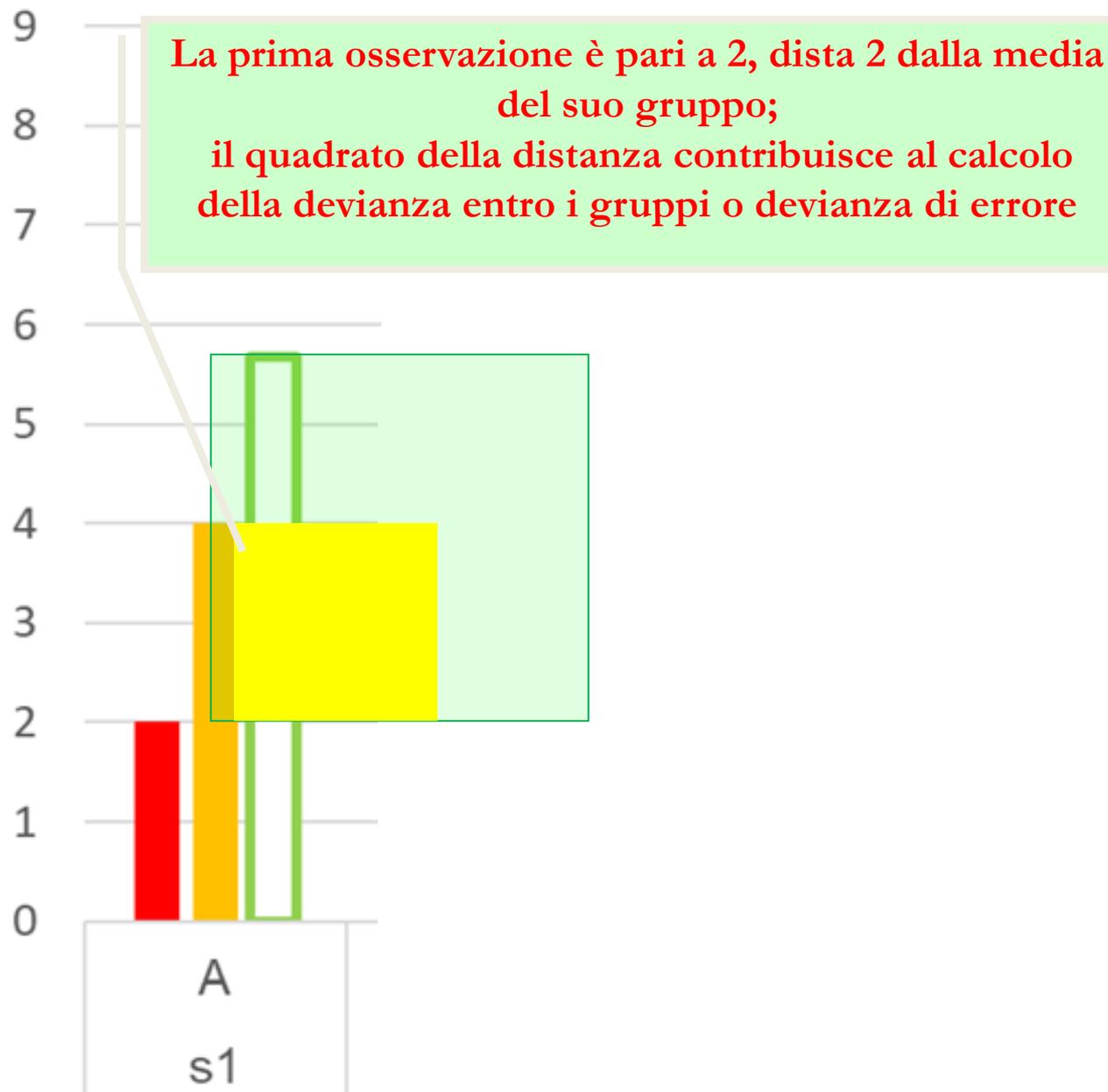
Tre tipi di distanze

9
8
7
6
5
4
3
2
1
0

A
s1

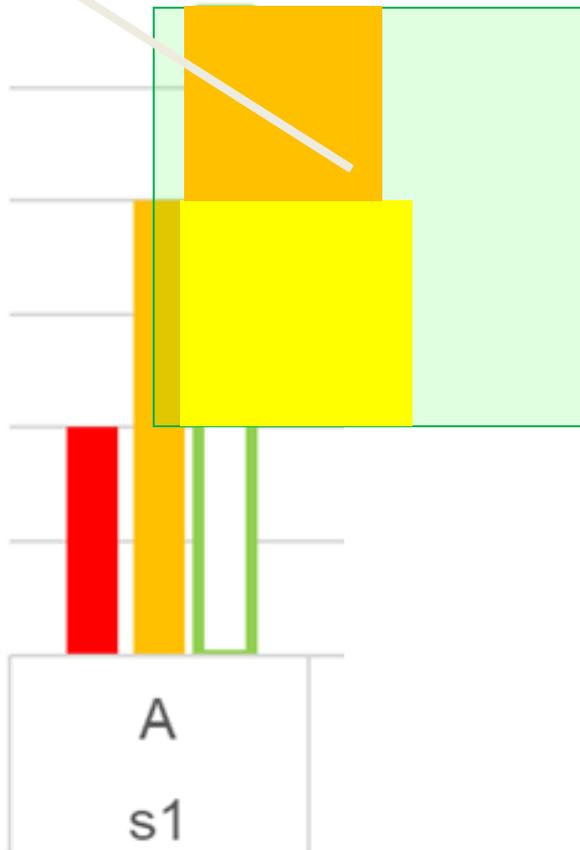


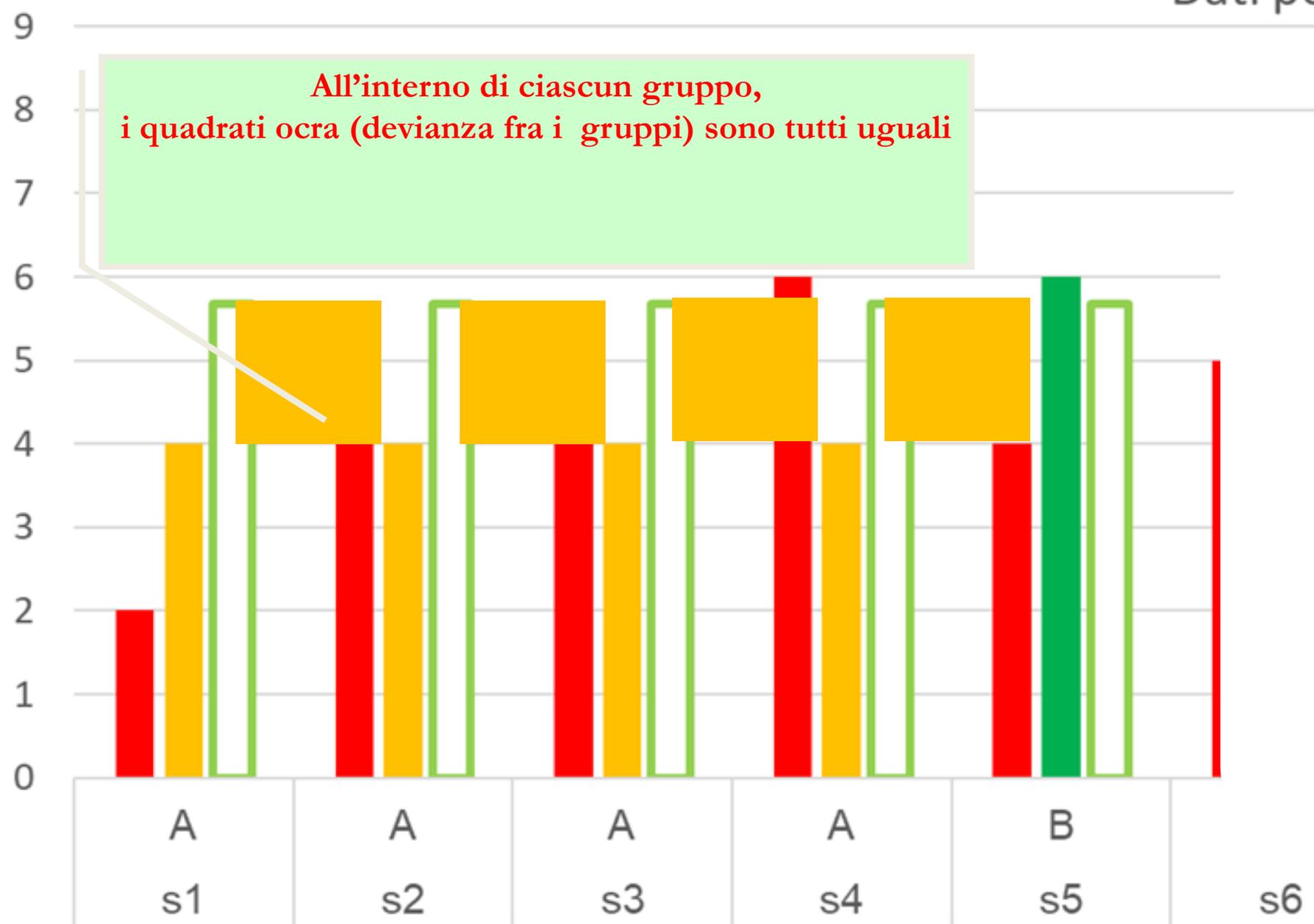




9
8
7
6
5
4
3
2
1
0

La prima osservazione ha un gruppo la cui media è 4;
il quadrato della distanza contribuisce al calcolo della
devianza fra i gruppi





All'interno di ciascun gruppo,
i quadrati ocra (devianza fra i gruppi) sono tutti uguali

Concludendo...

- Se le k medie sono simili, la variabilità fra i k gruppi è bassa, la varianza della popolazione è stimata in modo corretto, (tenuto conto della variabilità stocastica), il rapporto F è vicino all'unità e si conclude con l'accettazione di H_0 .
- Se c'è molta variabilità fra i k gruppi, la variabilità **fra** i gruppi è elevata, la varianza della popolazione è sovrastimata, il rapporto F è molto più grande dell'unità, il test statistico di F dà valori di probabilità molto bassi
- Se la probabilità di ottenere il valore F calcolato è molto bassa, si conclude con il **rifiuto** dell'ipotesi di nullità di differenze, per accettare l'ipotesi alternativa: almeno un gruppo proviene da una popolazione diversa, ossia con medie diverse

ANOVA per due gruppi?

- Il test dell'ANOVA dà gli stessi risultati della *t* di Student: infatti il rapporto F è il quadrato della *t*.