

2 Confronto fra medie

Statistica9 - 9/11/2015

tramite L'Analisi della Varianza

Consiste nell'attribuire una quota della Varianza totale di tutti i dati a sorgenti di variazione differenti e ben determinate e valutarne i rapporti di grandezza

Le varianze parziali così trovate vengono confrontate fra di loro per mezzo di un test denominato **test F**

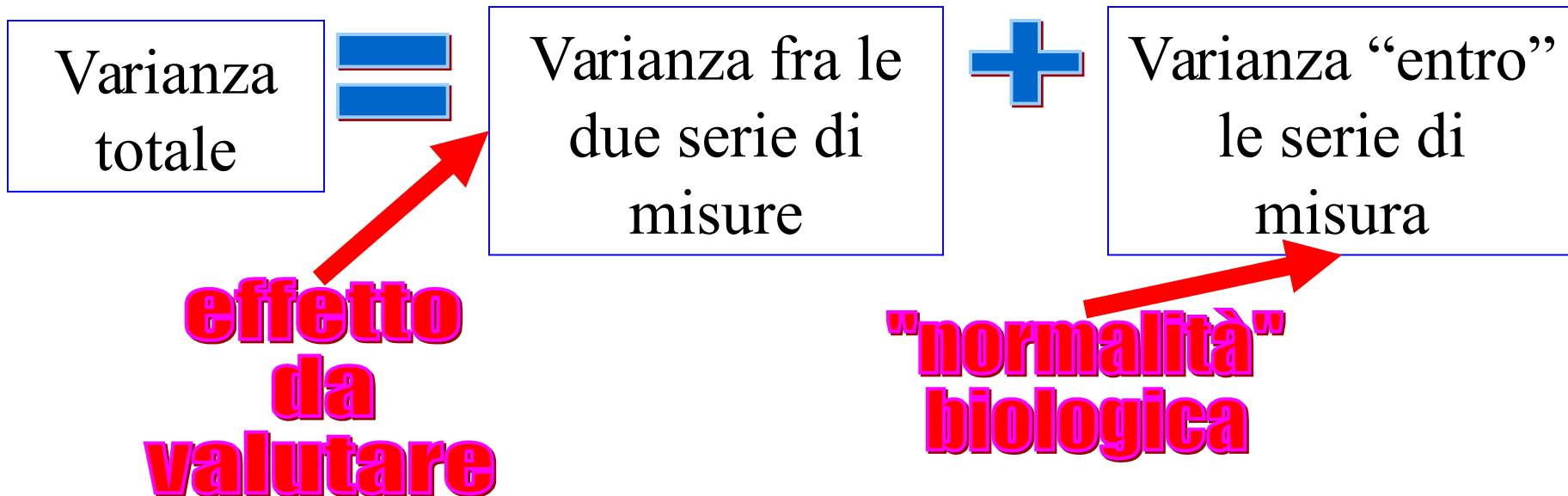
Il test F consiste essenzialmente nel confrontare (rapportare) la varianza aggiuntiva, dovuta agli effetti indotti (= trattamenti o fonti di variazione), con la varianza “naturale o di fondo” del tipo di misure che costituiscono le serie le cui medie devono essere analizzate.

se la varianza aggiuntiva, dovuta agli effetti indotti (= trattamenti o fonti di variazione) è “molto” superiore alla varianza “naturale o di fondo” evidentemente ciò non può essere dovuto al caso...

Il test



Due serie di misure



- Le deviazioni standard e le varianze non possiedono le proprietà additive (non le posso usare direttamente);
- le somme dei quadrati degli scarti ed i gradi di libertà possiedono le proprietà additive (**devo usare loro**).

quindi:

SS-TOT
somma dei
quadrati
degli scarti
totale

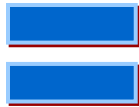


SS-tr somma
dei quadrati
degli scarti **fra**
le due serie di
misure



SS-er somma
dei quadrati
degli scarti
entro le due
serie di misure

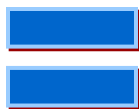
g.l. totali



g.l. delle medie
(es. se due serie, g.l.=1)



g.l. della prima
serie + g.l. della
seconda serie



Varianza fra le due serie di misure

Varianza entro le due serie di misure

Ipotesi zero H_0

cioè

**caratteristica
"normalità"
biologica**

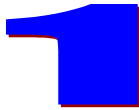
le 2 popolazioni A e B sono uguali cioè $\mu_a = \mu_b$

La varianza fra gruppi e la varianza entro gruppi sono due stime della stessa varianza che è la varianza della popolazione

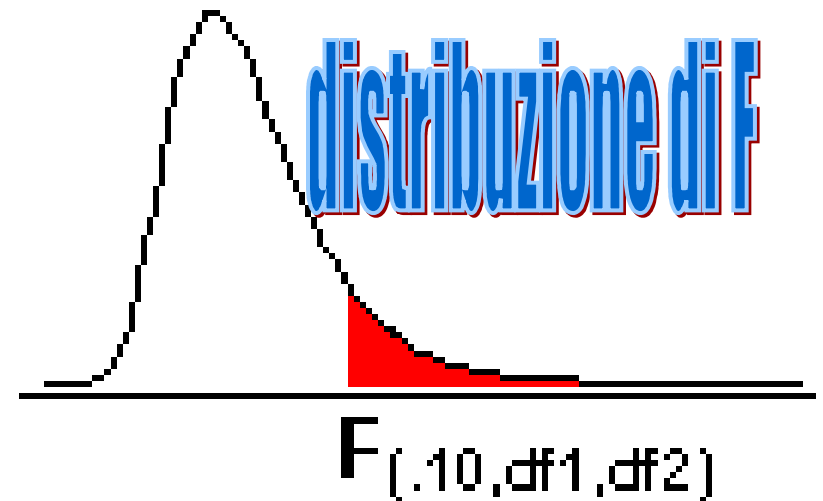
**ne consegue:
Per la definizione**

Varianza fra
le due serie
di misure

=



Varianza entro
le due serie di
misure



Questo rapporto fra varianza oscillerà intorno al valore 1 secondo una distribuzione o meglio una famiglia di distribuzioni - **il valore di F differisce secondo i gradi di libertà del numeratore e secondo i gradi di libertà del denominatore** - i cui andamenti sono stati definita da R. Fisher e riassunti in apposite tabelle.

Possiamo quindi calcolare il valore di F e controllare se il valore trovato sperimentalmente è compreso o supera il valore critico che viene superato solo nel 5% o nel 1% dei casi.

Se il valore trovato **SUPERA** il valore tabulato. Il valore trovato è **incompatibile** con l'**ipotesi \emptyset** e la **differenza** fra i valori è **statisticamente significativa** (o **altamente significativa**)

Se il valore trovato **NON SUPERA** il valore tabulato. Il valore trovato **non** è **incompatibile** con l'**ipotesi \emptyset** e la **differenza** fra i valori è da considerarsi **casuale**.

probabilità =0,05 di un valore più elevato di F.

g.l.	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25	30
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88	243.91	245.36	246.46	247.32	248.01	249.26	250.1
2	18.513	19	19.164	19.247	19.296	19.33	19.353	19.371	19.385	19.396	19.413	19.424	19.433	19.44	19.446	19.456	19.462
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.786	8.745	8.715	8.692	8.675	8.66	8.634	8.617
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964	5.912	5.873	5.844	5.821	5.803	5.769	5.746
5	6.608	5.786	5.409	5.192	5.05	4.95	4.876	4.818	4.772	4.735	4.678	4.636	4.604	4.579	4.558	4.521	4.496
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.06	4	3.956	3.922	3.896	3.874	3.835	3.808
7	5.591	4.737	4.347	4.12	3.972	3.866	3.787	3.726	3.677	3.637	3.575	3.529	3.494	3.467	3.445	3.404	3.376
8	5.318	4.459	4.066	3.838	3.687	3.581	3.5	3.438	3.388	3.347	3.284	3.237	3.202	3.173	3.15	3.108	3.079
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.23	3.179	3.137	3.073	3.025	2.989	2.96	2.936	2.893	2.864
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.02	2.978	2.913	2.865	2.828	2.798	2.774	2.73	2.7
12	4.747	3.885	3.49	3.259	3.106	2.996	2.913	2.849	2.796	2.753	2.687	2.637	2.599	2.568	2.544	2.498	2.466
14	4.6	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602	2.534	2.484	2.445	2.413	2.388	2.341	2.308
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494	2.425	2.373	2.333	2.302	2.276	2.227	2.194
18	4.414	3.555	3.16	2.928	2.773	2.661	2.577	2.51	2.456	2.412	2.342	2.29	2.25	2.217	2.191	2.141	2.107
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348	2.278	2.225	2.184	2.151	2.124	2.074	2.039
25	4.242	3.385	2.991	2.759	2.603	2.49	2.405	2.337	2.282	2.236	2.165	2.111	2.069	2.035	2.007	1.955	1.919
30	4.171	3.316	2.922	2.69	2.534	2.421	2.334	2.266	2.211	2.165	2.092	2.037	1.995	1.96	1.932	1.878	1.841
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.18	2.124	2.077	2.003	1.948	1.904	1.868	1.839	1.783	1.744
50	4.034	3.183	2.79	2.557	2.4	2.286	2.199	2.13	2.073	2.026	1.952	1.895	1.85	1.814	1.784	1.727	1.687
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927	1.85	1.792	1.746	1.708	1.676	1.616	1.573
150	3.904	3.056	2.665	2.432	2.274	2.16	2.071	2.001	1.943	1.894	1.817	1.758	1.711	1.673	1.641	1.58	1.535

Tavola **F** (0,01)

probabilità =0,01 di un valore più elevato di F.

g.l.	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20	25
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6143	6170	6192	6209	6240
2	98.5	99	99.17	99.25	99.3	99.33	99.36	99.37	99.39	99.4	99.42	99.43	99.44	99.44	99.45	99.46
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	27.05	26.92	26.83	26.75	26.69	26.58
4	21.2	18	16.69	15.98	15.52	15.21	14.98	14.8	14.66	14.55	14.37	14.25	14.15	14.08	14.02	13.91
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.888	9.77	9.68	9.61	9.553	9.449
6	13.75	10.92	9.78	9.148	8.746	8.466	8.26	8.102	7.976	7.874	7.718	7.605	7.519	7.451	7.396	7.296
7	12.25	9.547	8.451	7.847	7.46	7.191	6.993	6.84	6.719	6.62	6.469	6.359	6.275	6.209	6.155	6.058
8	11.26	8.649	7.591	7.006	6.632	6.371	6.178	6.029	5.911	5.814	5.667	5.559	5.477	5.412	5.359	5.263
9	10.56	8.022	6.992	6.422	6.057	5.802	5.613	5.467	5.351	5.257	5.111	5.005	4.924	4.86	4.808	4.713
10	10.04	7.559	6.552	5.994	5.636	5.386	5.2	5.057	4.942	4.849	4.706	4.601	4.52	4.457	4.405	4.311
12	9.33	6.927	5.953	5.412	5.064	4.821	4.64	4.499	4.388	4.296	4.155	4.052	3.972	3.909	3.858	3.765
14	8.862	6.515	5.564	5.035	4.695	4.456	4.278	4.14	4.03	3.939	3.8	3.698	3.619	3.556	3.505	3.412
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.89	3.78	3.691	3.553	3.451	3.372	3.31	3.259	3.165
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.597	3.508	3.371	3.269	3.19	3.128	3.077	2.983
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.457	3.368	3.231	3.13	3.051	2.989	2.938	2.843
25	7.77	5.568	4.675	4.177	3.855	3.627	3.457	3.324	3.217	3.129	2.993	2.892	2.813	2.751	2.699	2.604
30	7.562	5.39	4.51	4.018	3.699	3.473	3.304	3.173	3.067	2.979	2.843	2.742	2.663	2.6	2.549	2.453
40	7.314	5.179	4.313	3.828	3.514	3.291	3.124	2.993	2.888	2.801	2.665	2.563	2.484	2.421	2.369	2.271
50	7.171	5.057	4.199	3.72	3.408	3.186	3.02	2.89	2.785	2.698	2.562	2.461	2.382	2.318	2.265	2.167
100	6.895	4.824	3.984	3.513	3.206	2.988	2.823	2.694	2.59	2.503	2.368	2.265	2.185	2.12	2.067	1.965

Sorgenti di variazione	Somme dei quadrati degli scarti SS	gradi di libertà gl o df	Varianze MS	Rapporti F
Fra le serie o fra gruppi TRATTAMENTI	SS_{a-b}	n.gruppi-1	$(SS_{a-b}) / (n.\text{gruppi}-1)$	
entro le serie ERRORE	$SS_a + SS_b$	$(n-1)di_a + (n-1)di_b$	$(SS_a + SS_b) / ((n-1)di_a + (n-1)di_b)$	MStrattamenti / MErrore
TOTALE	SS_{ab}	gl_{ab}		

TRATTAMENTI:
variazione che stiamo cercando di determinare e che è dovuta ad un fattore conosciuto

Comunemente chiamato **ERRORE:** variazioni non controllate dovute al caso (variabilità caratteristica di tutti i parametri biologici).

In sostanza la varianza che influenza i risultati delle misure viene divisa in più parti:

UNA O PIÙ PARTI
TRATTAMENTI

FATTORE “CONOSCIUTO”:
costituita cioè da un fattore di cui si vuole verificare l’azione ed è stato causato e
CONTROLLATO sperimentalmente.

L’ALTRA PARTE o
ERRORE

FATTORE SCONOSCIUTO: tutto
quello che resta. Costituita cioè da tutti i
fattori non controllati - casuali.

Si verifica poi se la variazione dovuta ai fattori controllati è spiegabile con (rientra nel-) l’ampiezza della variazione dovuta ai fattori non controllati.

esempio numerico:

	A	B
	2	4
	3	5
	4	6
n	3	3
media	3	5
somma	9	15

TOTALE

media: $24/6=4$

$$4-2 = +2^2 = 4$$

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

$$4-4 = +0^2 = 0$$

$$4-4 = +0^2 = 0$$

$$4-5 = -1^2 = 1$$

$$4-6 = -2^2 = 4$$

$$SS = \frac{10}{5} = 2$$

$$\text{g.l. } (6-1)$$

ENTRO

media $a = 3, b = 5$

$$3-2 = +1^2 = 1$$

$$3-3 = +0^2 = 0$$

$$3-4 = -1^2 = 1$$

$$5-4 = +1^2 = 1$$

$$5-5 = +0^2 = 0$$

$$5-6 = -1^2 = 1$$

$$SS = \frac{4}{4} = 1$$

$$\text{g.l. } (2+2)$$

TRA

media $(3+5)/2=4$

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

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

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

$$4-5 = -1^2 = 1$$

$$4-5 = -1^2 = 1$$

$$4-5 = -1^2 = 1$$

$$SS = \frac{6}{1} = 6$$

$$\text{g.l. } (2-1)$$

tesso esempio numerico con T.C.:

	A	B
	2	4
	3	5
	4	6
n	3	3
media	3	5
somma	9	15

TOTALE

$$2^2 = 4$$

$$3^2 = 9$$

$$4^2 = 16$$

$$4^2 = 16$$

$$5^2 = 25$$

$$6^2 = 36$$

$$\text{tot} = 106$$

$$9 + 15 = 24^2 = 576$$

$$576/6 = 96$$

$$\text{SS} = 106 - 96 = \frac{10}{5}$$

$$\text{g.l.} (6-1) =$$

ENTRO

$$2^2 = 4$$

$$3^2 = 9$$

$$4^2 = 16$$

$$\text{tot} = 29$$

$$9^2 = 81$$

$$81/3 = 27$$

$$4^2 = 16$$

$$5^2 = 25$$

$$6^2 = 36$$

$$\text{tot} = 77$$

$$15^2 = 225$$

$$225/3 = 75$$

$$\text{SS} = 29 - 27 + 77 - 75 = \frac{4}{4}$$

$$\text{g.l.} (3-1+3-1) =$$

TRA

$$9^2 = 81/3 = 27$$

$$15^2 = 225/3 = 75$$

$$102$$

$$9 + 15 = 24^2 = 576$$

$$576/6 = 96$$

$$\text{SS} = 102 - 96 = \frac{6}{1}$$

$$\text{g.l.} (2-1) =$$

meglio!

ENTRO

$$\text{SS} = 10 - 6 = \frac{4}{4}$$

$$\text{g.l.} (5-1) =$$

12

TRA

$$3^2 + 3^2 + 3^2 = 9^2 / 3 = 81/3 = 27$$

$$5^2 + 5^2 + 5^2 = 15^2 / 3 = 225/3 = 75$$

$$27 + 75 = \mathbf{102}$$

$$\mathbf{TC} = 9 + 15 = 24^2 = 576 ; 576/6 = \mathbf{96}$$

$$\mathbf{SS} = 102 - 96 = \underline{6}$$

$$\mathbf{g.l.} (2-1) = 1$$

CIOE'

$$\text{media}^2 + \text{media}^2 + \text{media}^2 = \text{somma}^2 / 3 = 81/3 = 27$$

$$\text{media}^2 + \text{media}^2 + \text{media}^2 = \text{somma}^2 / 3 = 225/3 = 75$$

generalizzando:

$$x^2 + x^2 + x^2 \text{ "n" volte} = n \times x^2 = \frac{n \times n \times x^2}{n} = \frac{(n \times x)^2}{n}$$

$$\text{Cioè} \quad \frac{\text{somma}^2}{n}$$

Ricorda: 

Somma quadrati scarti è uguale a dire somma dei quadrati dei dati meno somma dei dati al quadrato diviso la numerosità

Nel caso del calcolo della SS fra le serie (trattamenti)

la somma dei quadrati dei dati sarebbe: $3^2 + 3^2 + 3^2 + 5^2 + 5^2 + 5^2 = 102$

(il termine di correzione rimane **96** cioè $(2+3+4+4+5+6)^2/6$)

Che è lo stesso che calcolare $\frac{(\text{somma di A})^2}{n \text{ di A}} + \frac{(\text{somma di B})^2}{n \text{ di B}} - TC$

infatti:

molto più semplice e rapido!

$$102 = 3^2 + 3^2 + 3^2 + 5^2 + 5^2 + 5^2 = 9+9+9+25+25+25 = \frac{(2+3+4)^2}{3} + \frac{(4+5+6)^2}{3} =$$

$$\text{media-di-A}^2 + \text{media-di-A}^2 + \text{media-di-A}^2 + \text{media-di-B}^2 + \text{media-di-B}^2 + \text{media-di-B}^2 = \frac{(\text{somma di A})^2}{n} + \frac{(\text{somma di B})^2}{n} =$$

ricordando la definizione di media aritmetica come somma dei dati diviso il numero degli stessi.

Sorgenti di variazione	Somme dei quadrati degli scarti SS	gradi di libertà gl o df	Varianze MS	Rapporti F
Fra le serie o fra gruppi TRATTAMENTI	6	1	6	6
entro le serie ERRORE	4	4	1	
TOTALE	10	5	2	

Tavola

F
(0,05)

gl -1 e 4
7,709

$6 < 7,709$

cioè

il valore trovato **NON SUPERA** il valore tabulato. Il valore trovato **non è incompatibile** con l'ipotesi \emptyset e la **differenza** fra i valori è da considerarsi **casuale**.

ES:

La media di A risulta statisticamente differente da quella di B? $A \neq B$ per $p < 0,05$ (oppure no)

PESI DI A PESI DI B A+B

59 62

45 46

57 79

49 79

20 82

54

n = 6 5 11

tot = 284 348 632

media = 47.333333 69.6 57.45

**Confronto
fra medie
tramite**

**2 L'Analisi della
Varianza**

ENTRO		²	²	g.l.
-11,66667	7,6	136,11111	57,76	
2,3333333	23,6	5,4444444	556,96	
-9,666667	-9,4	93,444444	88,36	
-1,666667	-9,4	2,7777778	88,36	
27,3333333	-12,4	747,11111	153,76	
-6,666667		44,444444		
somma		1974,5333		9

TOTALE		²	²	g.l.
-1,545455	-4,545455	2,3884298	20,661157	
12,454545	11,454545	155,1157	131,20661	
0,4545455	-21,54545	0,2066116	464,20661	
8,4545455	-21,54545	71,479339	464,20661	
37,454545	-24,54545	1402,843	602,47934	
3,4545455	57,454545	11,933884		
somma		3326,7273		10

TRA		²	²	g.l.
10,121212	-12,14545	102,43893	147,51207	
10,121212	-12,14545	102,43893	147,51207	
10,121212	-12,14545	102,43893	147,51207	
10,121212	-12,14545	102,43893	147,51207	
10,121212	-12,14545	102,43893	147,51207	
10,121212		102,43893		
somma		1352,1939		1

Sorgenti di variazione	Somme dei quadrati degli scarti SS	gradi di libertà gl o df	Varianze MS	Rapporti F
Fra le serie o fra gruppi TRATTAMENTI	1352,194	1	1352,194	
entro le serie ERRORE	1974,533	9	219,393	6,163
TOTALE	3326,727	10		

Tavola (0,05)

F

gl 1 9
5,117

6,163 > 5,117

cioè

solo significativa

(0,01)

gl 1 9
10,56

6,163 < 10,56

Se il valore trovato **SUPERA** il valore tabulato. Il valore trovato è **incompatibile** con l'ipotesi \emptyset e la **differenza** fra i valori è **statisticamente significativa**

PESI DI A PESI DI B

n =	6	5
media =	47.3 a	69.6 b
d.s. =	14.35	15.37

Nota: lettere diverse indicano differenze significative per $p < 0,05$

e.s. = 5,858363 6,873673

stesso esempio usando il T.C.:

A	B	A ²	B ²
59	62	3481	3844
45	46	2025	2116
57	79	3249	6241
49	79	2401	6241
20	82	400	6724
54		2916	

“complicato”, più rapido e più semplice calcolarla per differenza

	TOTALE		
n	6	5	11
media	47,33333	69,6	57,45455
somma	284	348	632
quadrat	80656	121104	399424

$$\begin{aligned}
 & 14472 - \frac{80656}{6} = 1029,33 \\
 & 25166 - \frac{121104}{5} = 945,2 \\
 & 14472 - \frac{80656}{6} + 25166 - \frac{121104}{5} = 1974,53 \\
 & 39638 - \frac{399424}{11} = 3326,73
 \end{aligned}$$

Sorgenti di variazione	Somme dei quadrati degli scarti SS	gradi di libertà gl o df	Varianze MS	Rapporti F
Fra le serie o fra gruppi TRATTAMENTI	1352,194	1	1352,194	
entro le serie ERRORE	1974,533	9	219,393	6,163
TOTALE	3326,727	10		

**Usare sempre il metodo del
termine di correzione
per il calcolo delle
somme dei quadrati degli scarti, varianze ecc.!**

**La “formula” da applicare è meno intuitiva e più complicata
ma i calcoli non comportano cifre decimali da elevare al
quadrato e sono molto più rapidi e semplici!**