CORSO DI STATISTICA (parte 1) - ESERCITAZIONE 4

Dott.ssa Antonella Costanzo

a.costanzo@unicas.it

<u>Verifica delle proprietà della media aritmetica, indici di variabilità per caratteri quantitativi</u>

Il seguente data set riporta la rilevazione di alcuni caratteri su un collettivo di 20 studenti.

Studente	Sesso	Età	Red	Istituto di provenienza	Voto al diploma	Statura (cm)	Colore occhi	Voto esame	Giud. sul corso
1	M	22	0,7	ITC	96	173	Nero	26	Pessimo
2	F	20	0,2	Liceo Classico	92	168	Marrone	26	Ottimo
3	F	30	1,6	Liceo Classico	90	165	Marrone	30	Buono
4	M	22	2,5	Liceo Scient	85	180	Nero	25	Buono
5	F	26	3,2	ITI	100	163	Azzurro	30	Pessimo
6	F	20	0,5	ITC	74	160	Nero	24	Pessimo
7	M	26	4,2	Liceo Scient	60	177	Marrone	20	Suff
8	M	30	1,3	ITC	76	164	Verde	18	Ottimo
9	F	27	1,2	Liceo Scient	100	158	Azzurro	29	Ottimo
10	F	25	1,7	ITI	95	170	Nero	25	Pessimo
11	F	25	1,9	ITI	85	167	Nero	25	Buono
12	M	22	0,7	ITC	97	159	Marrone	27	Buono
13	F	21	0,4	Liceo Classico	65	174	Azzurro	21	Ottimo
14	F	24	1,8	Liceo Scient	70	164	Verde	30	Suff
15	M	20	1,9	Liceo Scient	80	177	Nero	28	Suff
16	F	21	3,2	Liceo Classico	93	172	Nero	27	Pessimo
17	F	27	2,1	ITC	100	166	Marrone	26	Suff
18	F	22	0,1	ITI	84	160	Marrone	24	Buono
19	M	23	1,6	Liceo Scient	92	170	Azzurro	27	Ottimo
20	F	23	2,2	Liceo Scient	73	184	Verde	23	Buono

Esercizio 1. Le proprietà della media aritmetica

- Linearità (invarianza a trasformazioni lineari)
- Internalità (la media è sempre compresa tra un minimo e un massimo)
- Baricentricità (la somma degli scarti dalla media è zero)
- Associatività (se un collettivo viene diviso in G sottoinsiemi disgiunti, allora la media aritmetica generale si può ottenere come media ponderata delle medie dei sottoinsiemi con pesi uguali alle loro numerosità)
- Criterio di minimo della media (la media è quel valore che minimizza la somma degli scarti al quadrato)

Verificare la **proprietà associativa** della media aritmetica del voto all'esame rispetto al sesso degli studenti.

Nel caso specifico dobbiamo verificare che la media generale del Voto all'esame è uguale alla media ponderata delle medie di gruppo (maschi/femmine) con pesi dati dalla numerosità dei gruppi. A tale scopo, costruiamo le distribuzioni condizionate del carattere Voto all'esame rispetto alle modalità del carattere genere (maschio/femmina); quindi calcoliamo le medie di gruppo e la media generale.

Distribuzioni condizionate del carattere Voto all'esame per maschi e femmine

Voto all'esame	Femmine	Maschi
18	0	1
20	0	1
21	1	0
23	1	0
24	2	0
25	2	1
26	2	1
27	1	2
28	0	1
29	1	0
30	3	0
totale	13	7

$$\mu_{Voto\ Esame} = \frac{18*1+20*1+\cdots}{20} = 25.55$$

$$\mu_{Voto\ Esame|Maschi} = \frac{18*1+20*1+21*0+23*0+\cdots}{7} = 24.43$$

$$\mu_{Voto\ Esame|Femmine} = \frac{18*0+20*0+21*1+23*1+\cdots}{13} = 26.15$$

$$\mu_{Voto\ Esame} = \frac{24.43*7+26.15*13}{20} = 25.55$$

Facendo riferimento nuovamente alla distribuzione della variabile reddito, verificare la **proprietà della baricentricità** della media aritmetica

Reddito (classi in Euro)	c_i	f
[0.1, 1)	0.55	0.3
[1,1.9)	1.45	0.3
[1.9, 4.2]	3.05	0.4
Totale (n)		1

Per dati organizzati in classi, dobbiamo dimostrare che vale la seguente:

$$\sum_{i=1}^{n}(c_i - \mu)f_i = 0$$

$$\mu = 0.55(0.3) + 1.45(0.3) + 3.05(0.4) = 1.82$$

$$[0.55 - 1.82]0.3 + [1.45 - 1.82]0.3 + [3.05 - 1.82]0.4 = -0.381 - 0.111 + 0.492 = 0$$

La **variabilità** è l'attitudine di un fenomeno ad assumere modalità differenti. Esistono diverse misure di variabilità di una distribuzione di frequenza. In particolare, vedremo per le variabili quantitative:

- **Variabilità rispetto alle singole unità (variabilità reciproca):** misurare la diversità tra le singole unità statistiche, ossia quanto ciascuna unità statistica è "diversa" rispetto a tutte le altre (differenza semplice media, differenza quadratica media)
- **Variabilità rispetto a particolari posizioni delle unità nella distribuzione:** misurare la diversità fra le modalità di due particolari unità della distribuzione o fra due quantili (campo di variazione, CVI)
- **Variabilità rispetto ad un "centro":** controllare se le singole unità statistiche presentano modalità più o meno stabili rispetto ad un indice di posizione che è rappresentativo dell'intera distribuzione di frequenza (scostamento semplice medio dalla media, scostamento semplice medio dalla mediana, scostamento quadratico medio dalla media (scarto quadratico medio), varianza.

Esercizio 2. Indici di mutua variabilità o variabilità reciproca

Con riferimento alla variabile reddito riportata in tabella, calcolare la differenza semplice media e la differenza media quadratica.

Reddito (classi)	c_i	n_i	fi
[0.1, 1)	0.55	6	0.3
[1,1.9)	1.45	6	0.3
[1.9, 4.2]	3.05	8	0.4
Totale (n)		20	1

Come già accennato, la variabilità di un fenomeno può essere studiata anche in termini di differenza di ciascun dato da tutti gli altri. Si definisce differenza media di una distribuzione di dati una media calcolata sulle differenze fra ciascun dato e tutti gli altri.

Tutte le differenze possibili compresa la differenza dell'i-esimo valore da se stesso è una matrice nxn. Nel nostro caso:

$c_i - c_j$

c_i/c_j	**0.55**	**1.45**	**3.05**
0.55	0	-0.9	-2.5
1.45	0.9	0	-1.6
3.05	2.5	1.6	0

Tutte le possibili differenze sono n^2. Tuttavia, poiché i termini della diagonale principale sono nulli le differenze non nulle sono $n^2 - n = n(n-1)$. Una volta individuate tutte le possibili distanze di ogni dato da tutti gli altri ne faccio la media e ottengo una misura di variabilità che prende appunto il nome di differenza semplice media:

Differenza semplice media con ripetizione $\Delta_r = \frac{\sum_{i=1}^{h}\sum_{j=1}^{k}|c_i - c_j| n_i * n_j}{n^2}$ quando si considerano tutte le n^2 differenze (comprese quelle nulle)

Differenza semplice media senza ripetizione $\Delta = \frac{\sum_{i=1}^{h}\sum_{j=1}^{k}|c_i - c_j| n_i * n_j}{n(n-1)}$ quando si considerano solo le $n(n-1)$ differenze escludendo i termini nulli della diagonale principale

Per agevolare i calcoli si può definire anche la matrice dei prodotti $n_i * n_j$:

$n_i * n_j$

n_i/n_j	**6**	6	8
6	36	36	48
6	36	36	48
8	54	54	64

A questo punto il numeratore della differenza media semplice (senza ripetizione) si ottiene moltiplicando elemento per elemento le due tabelle precedenti e sommando i prodotti ottenuti:

$|c_i - c_j|$

c_i/c_j	**0.55**	**1.45**	**3.05**
0.55	0	0.9	2.5
1.45	0.9	0	1.6
3.05	2.5	1.6	0

$n_i * n_j$

n_i/n_j	**6**	6	8
6	36	36	48
6	36	36	48
8	48	48	64

$$\Delta = \frac{0.9*36+2.5*48+0.9*36+1.6*48+2.5*48+1.6*48}{20(20-1)} = 1.21$$

Nota: il valore minimo di Δ è 0 e si verifica quando i dati coincidono tra di loro per cui tutte le differenze sono nulle.

Se anziché considerare il valore assoluto delle differenze semplici considero le differenze al quadrato e ne faccio la media ottengo una misura della differenza quadratica media:

Differenza media quadratica (senza ripetizione): $\Delta_2 = \sqrt{\frac{\sum_{i=1}^{h}\sum_{j=1}^{k}(c_i-c_j)^2*n_in_j}{n(n-1)}}$ ossia la media quadratica delle differenze non nulle. Sfruttando le tabelle costruite sopra per la differenza semplice media, per comodità calcoliamo prima il radicando e poi il risultato finale:

Radicando:

$$\frac{(-0.9)^2*36+(-2.5)^2*48+(0.9)^2*36+(-1.6)^2*48+(2.5)^2*48+(1.6)^2*48}{20(20-1)} =$$

$$= \frac{29.16+300+29.16+122.88+300+122.88}{20(20-1)} = 2.38$$

Quindi:

$$\Delta_2 = \sqrt{2.38} = 1.54$$

Esercizio 3. Variabilità rispetto ad un centro

Consideriamo sempre la variabile reddito suddivisa in classi. Sapendo che la media è pari a $\bar{x}$ =1.82 e la mediana è pari a 1.60 (migliaia di euro), calcolare: lo scostamento semplice medio dalla media e lo scostamento semplice medio dalla mediana, lo scostamento quadratico medio dalla media (scarto quadratico medio), la devianza e la varianza.

Reddito (classi)	c_i	n_i	fi	$c_i-\bar{x}$	$\lvert ci-\bar{x}\rvert * n_i$	$(c_i-\bar{x})^2$	$(c_i-\bar{x})^2*n_i$	$\lvert c_i-Me\rvert * n_i$
[0.1, 1)	0.55	6	0.3	-1.27	7.62	1.613	9.678	6.3
[1,1.9)	1.45	6	0.3	-0.37	2.22	0.137	0.822	0.9
[1.9, 4.2]	3.05	8	0.4	1.23	9.84	1.513	12.104	11.6
Totale (n)		20	1		19.68		22.604	18.80

Scostamento semplice medio dalla media: $S_{\bar{x}} = \frac{\sum_{i=1}^{n}|c_i-\bar{x}|*n_i}{n} = \frac{19.68}{20} = 0.984$

Tanto più piccolo è lo scarto semplice medio tanto più i valori si addensano attorno alla media aritmetica. Nota: lo scostamento semplice medio dalla media è una misura che non richiede l'ordinamento dei dati, necessario invece per il calcolo della mediana:

Scostamento semplice medio dalla mediana: $S_{Me} = \frac{\sum_{i=1}^{n}|c_i - \boldsymbol{Me}| * n_i}{n} = \frac{18.80}{20} = 0.94$

Lo scarto semplice medio dalla mediana gode dell'importante proprietà di essere il più piccolo fra tutti gli scarti medi della variabile da qualsiasi valore medio.

Scostamento quadratico medio dalla media (o scarto quadratico medio):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(c_i - \bar{x})^2 n_i}{n}} = \sqrt{\frac{22.604}{20}} = 1.06$$

Varianza: $\sigma^2 = \frac{22.604}{20} = 1.13$

La varianza rappresenta sostanzialmente lo scostamento (o scarto) di ciascun valore rispetto alla media espresso al quadrato dell'unità di misura. Lo scarto quadratico medio è infatti un indice di variabilità che non risente dell'unità di misura, invece la varianza è un indice assoluto che risente dell'unità di misura della variabile.

Devianza: numeratore della varianza=22.604

Si può dimostrare inoltre che la varianza può essere calcolata alternativamente come: $\sigma_x^2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} - \bar{x}^2$ ossia la varianza è pari alla media aritmetica dei quadrati meno il quadrato della media aritmetica. Nel nostro caso siccome si hanno dati in classi: $\sigma_x^2 = \frac{\sum_{i=1}^{n} c_i^2 n_i}{n} - \bar{x}^2$

La varianza assume valore minimo (zero) quando tutte le modalità sono uguali tra loro e aumenta all'aumentare della differenza tra i valori osservati. Il massimo può essere infinito perché gli scarti possono essere infinitamente grandi (ovvero le modalità infinitamente lontane dalla media aritmetica).

Esercizio 4. Confronti in termini di variabilità

Tutti gli studenti intervistati sono fuori sede. Siamo interessati a fare un confronto in termini di variabilità tra la variabile reddito percepito e la superficie della casa affittata agli studenti per il corrente anno accademico. La distribuzione della variabile Superficie (in mt quadri) è riportata nella tabella seguente:

Superficie (classi)	c_i	c_i^2	n_i	f_i	F_i
[20,30)	25	625	11	0.55	0.55
[30,45)	37.5	1406.25	3	0.15	0.70
[45,60)	52.5	2756.25	4	0.20	0.90
[60, 90]	75	5625	2	0.10	1
			20	1	

Media $\bar{x} = 37.38$

Me=29.09

Calcolare gli indici di variabilità che permettano il confronto tra la variabilità delle due variabili.

Indici di variabilità per il carattere Superficie

- Scostamento semplice medio dalla media: 13.61
- Scostamento semplice medio dalla mediana: 12.78
- Varianza:271.18
- Scarto quadratico medio:16.47
- Differenza semplice media: 17.41
- Differenza quadratica media: 23.90

Indici di variabilità della variabile Reddito precedentemente calcolati:

- Scostamento semplice medio dalla media: 0.984
- Scostamento semplice medio dalla mediana: 0.94
- Varianza:1.13
- Scarto quadratico medio:1.06
- Differenza semplice media: 1.21
- Differenza quadratica media: 1.54

Procedo quindi con il confronto tra la variabilità delle variabili Superficie e Reddito degli studenti.

Promemoria:

<u>Gli indici di variabilità assoluta non consentono di effettuare il confronto fra la variabilità di distribuzioni espresse in unità di misura diverse.</u> In particolare se due distribuzioni sono espresse in unità di misura non trasformabili non è possibile utilizzare gli indici assoluti per confrontare la variabilità delle distribuzioni.

Per effettuare un confronto in termini di variabilità tra le distribuzioni delle variabili espresse in unità di misura diverse abbiamo bisogno di indici relativi di variabilità:

Indici percentuali di variabilità o di variabilità relativa alla media

Siccome per gli indici di variabilità non si conosce il massimo che possono assumere è conveniente dividere il valore dell'indice per il corrispondente indice di posizione scelto per misurare la dispersione. Ad esempio: la versione relativa dello scostamento semplice medio dalla media e dello scostamento semplice medio dalla mediana si ottiene dividendo tali indici per la corrispondente media o la corrispondente mediana rispettivamente. Per cui avremo:

Scostamento semplice medio dalla media relativo:

$$S_x^{rel} = \frac{S_{\bar{x}}}{\bar{x}}$$

Scostamento semplice medio dalla mediana relativo:

$$S_{Me}^{rel} = \frac{S_{Me}}{Me}$$

Inoltre, dividendo lo scarto quadratico medio per la corrispondente media si ottiene il

Coefficiente di variazione :

$$CV = \frac{\sigma}{\bar{x}}$$

Il coefficiente di variazione è un indice relativo, indipendente dall'unità di misura e dall'ordine di grandezza della variabile. Pertanto può essere utilizzato per confrontare fenomeni espressi in unità di misura diverse o rilevati in momenti diversi. Ha il minimo uguale a zero e il massimo non definito, giacché varia al variare del tipo di distribuzione. *Note utili*: se la media è negativa, se ne considera il valore assoluto affinchè risulti positivo; se la media, in valore assoluto, risulta prossima a zero (per effetto di parziali compensazioni fra valori positivi e negativi), il CV può segnalare, in maniera errata, una variabilità molto elevata del fenomeno.

Per gli indici di variabilità reciproca nel caso generale (e quindi prescindendo dall'ipotesi di caratteri trasferibili) è possibile ottenere una versione relativa dividendo il valore degli indici sempre per la media:

Differenza semplice media relativa:

$$\Delta_{rel} = \frac{\Delta}{\bar{x}}$$

Differenza media quadratica relativa:

$$\Delta_2^{rel} = \frac{\Delta_2}{\bar{x}}$$

Tornando quindi alla soluzione dell'esercizio, gli indici di variabilità appropriati per poter effettuare un confronto tra la variabilità delle due variabili sono i seguenti:

Variabile Reddito:

scostamento semplice medio dalla media (relativo): $\frac{S_{\bar{x}}}{\bar{x}} = \frac{0.984}{1.82} = 0.54$

scostamento semplice medio dalla mediana (relativo): $\frac{S_{Me}}{Me} = \frac{0.94}{1.60} = 0.59$

differenza semplice media relativa $\Delta_{rel} = \frac{1.21}{1.82} = 0.66$

differenza quadratica media relativa: $\Delta_2^{rel} = \frac{1.54}{1.82} = 0.85$

coefficiente di variazione: CV=0.58

Variabile Superficie casa

scostamento semplice medio dalla media (relativo): $\frac{S_{\bar{x}}}{\bar{x}} = \frac{13.61}{37.38} = 0.36$

scostamento semplice medio dalla mediana (relativo): $\frac{S_{Me}}{Me} = \frac{12.78}{29.09} = 0.44$

differenza semplice media relativa: $\Delta_{rel} = \frac{17.41}{37.38} = 0.47$

differenza quadratica media relativa: $\Delta_2^{rel} = \frac{23.90}{37.38} = 0.64$

coefficiente di variazione: CV=0.44

Commenti: la distribuzione del reddito degli studenti è più variabile rispetto alla distribuzione della superficie della loro abitazione come si nota dai valori dei seguenti indici: CV, scostamento semplice medio dalla media e dalla mediana relativi. La variabile Reddito rispetto alla variabile Superficie presenta inoltre un maggior grado di disuguaglianza nei dati (maggiore dispersione di ciascun valore da tutti gli altri), indipendentemente da qualsiasi valore medio (o mediano) come si può evincere dal valore delle differenze medie (semplice e quadratica).

Le proprietà dello scarto quadratico medio:

Lo sqm gode di alcune importanti proprietà:

1. sqm di X è sempre un numero non negativo.
2. Se X è una costante *b*, allora sqm di X è zero $\sigma(b) = 0$. Gli indici di variabilità infatti dovrebbero sempre annullarsi in caso di variabilità nulla.
3. Se alla variabile X si aggiunge una costante b, sqm di X non cambia $\sigma(X) \equiv \sigma(X + b)$
4. Se si moltiplica la variabile X per una costante b ($X^* = bX$), si avrà che $\sigma(X^*) = |b|\sigma(X)$

Esercizio 5. Applicazione proprietà 3

La seguente tabella riporta la paga mensile percepita da 10 lavoratori dipendenti della ditta Alfa. Verificare la variabilità del reddito e trovare lo scarto quadratico medio del reddito nell'ipotesi che a ciascun individuo venga aumentato lo stipendio di 150 euro.

Lavoratori	**X=Stipendio (in euro)**
1	1200
2	1430
3	1350
4	1500
5	1100
6	1250
7	1000
8	1600
9	1400
10	1730

$\mu = 1356$

$$\sigma = \sqrt{\frac{(x_i - \mu)^\wedge 2}{n}} = \sqrt{50604.44} = 224.95$$

Svolgimento:

Lo scarto quadratico medio, è invariante per traslazione ovvero se viene aggiunta una costante a ciascun valore del carattere lo scarto quadratico medio non varia. Quindi:

$Y = a + X$ $\qquad\qquad\qquad$ $\sigma_y = \sigma_{a+X} = \sigma_x = 224.95$

Esercizio 6. Applicazione proprietà 4

A dicembre verrà rivisto il contratto di lavoro, quindi a partire dal 2013 per ciascun lavoratore è prevista un aumento del 15% della paga percepita. Determinare la varianza della nuova distribuzione dello stipendio percepito dai lavoratori.

Svolgimento:

Si tratta di una trasformazione (di scala) dello stipendio percepito dai lavoratori (X). In questo caso per la proprietà 4 è sufficiente moltiplicare X per il coefficiente b per avere l'indice di variabilità sulla scala trasformata.:

a=150 b=0.15 Y=a+bX

$\sigma_y = \sigma_{a+bX} = |b|\sigma_x = 0.15(224.95) = 33.74$

Nota: ricordando che la varianza è un indice di variabilità espresso al quadrato dell'unità di misura il coefficiente di traslazione b incide in modo quadratico su X.

Quindi, dati a e b costanti vale la seguente proprietà per la varianza: $\sigma^2_{a+bX} = b^2\sigma^2_X$ per cui si avrà che

$\sigma^2_{a+bX} = 0.15^2$(50604.44)=1138.60

Esercizio 7. La scomposizione della varianza

La seguente tabella riportata le informazioni congiunte relative al fatturato ottenuto nei due settori di attività dell'azienda Alfa, meccanico e chimico. La variabile Settore assume quindi due modalità: meccanico che indichiamo con M, e chimico che indichiamo con C. Determinare la varianza per ciascuna distribuzione parziale e successivamente verificare la scomposizione della varianza

X=Classi di fatturato|Y=Settore	c_i	n_i^M	n_i^C	**Totale (ni.)**
[0, 1)	0.5	15	25	**40**
[1, 3)	2	24	51	**75**
[3, 5)	4	85	67	**152**
[5, 10)	7.5	48	59	**107**
[10, 20)	15	40	31	**71**
[20, 40)	30	29	31	**60**
Totale (n.j)		**241**	**264**	**505**

Calcoliamo le medie e le varianze per ciascuna distribuzione parziale. Per agevolare i calcoli costruiamo la seguente tabella:

X=Classi di fatturato | Y=Settore	c_i	c_i^2	n_i^M	n_i^C	$c_i n_i^M$	$c_i n_i^C$	$c_i^2 n_i^M$	$c_i^2 n_i^C$
[0, 1)	0.5	0.25	15	25	7.5	12.5	3.75	6.25
[1, 3)	2	4	24	51	48	102	96	204
[3, 5)	4	16	85	67	340	268	1360	1072
[5, 10)	7.5	56.25	48	59	360	442.5	2700	3318.75
[10, 20)	15	225	40	31	600	465	9000	6975
[20, 40)	30	900	29	31	870	930	26100	27900
Totale			**241**	**264**	**2225.5**	**2220**	**39259.75**	**39476**

Medie condizionate:

$$\mu_{Fatt|M} = \bar{x}_{i1} = \frac{1}{241}(2225.5) = 9.234$$

$$\mu_{Fatt|C} = \bar{x}_{i2} = \frac{1}{264}(2220) = 8.409$$

Varianze condizionate:

$$\sigma^2_{Età|Mecc} = \frac{1}{241}\sum_{i=1}^{n} c_i^2 n_{i1} - (\mu_{Fatt|M})^2 = \frac{1}{241}[39259.75] - 9.234^2 = 77.64$$

$$\sigma^2_{Età|Chim} = \frac{1}{264}\sum_{i=1}^{n} c_i^2\, n_{i2} - (\mu_{Fatt|C})^2 = \frac{1}{264}[39476] - 8.409^2 = 78.81$$

A questo punto determino la media e la varianza dell'intera popolazione. Per agevolare il calcolo costruiamo la seguente tabella

c_i	$n_{i.}$	$c_i n_{i.}$	$c_i^2 n_{i.}$
0.5	40	20	10
2	75	150	300
4	152	608	2432
7.5	107	802.5	6018.75
15	71	1065	15975
30	60	1800	54000
Totale	505	4445.5	78735.75

Il reddito medio (o grande media) per l'intera popolazione è

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} c_i n_{i.} = \frac{4445.5}{505} = 8.803$$

la varianza della popolazione totale è quindi:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} c_i^2 n_i - \bar{x}^2 = \frac{78735.75}{505} - 8.803^2 = 78.422$$

Per verificare la proprietà di scomposizione della varianza dobbiamo calcolare:

- **Varianza interna ai gruppi (within) $\boldsymbol{\sigma_W^2}$:** descrive lo scostamento della singola osservazione dalla media del suo gruppo j-esimo
- **Varianza esterna ai gruppi (between) $\boldsymbol{\sigma_B^2}$:** descrive lo scostamento della media del gruppo j-esimo dalla grande media (media di tutti i valori).

E dimostrare che la varianza totale σ^2 può essere espressa come somma della varianza interna ai gruppi e la varianza esterna ai gruppi, ossia:

$$\boldsymbol{\sigma^2 = \sigma_W^2 + \sigma_B^2 = \frac{1}{n}\sum_{j=1}^{G} \sigma_j^2\, n_{.j} + \frac{1}{n}\sum_{j=1}^{G} (\bar{x}_j - \bar{x})^2\, n_{.j}}$$

$$\sigma_W^2 = \frac{1}{n}\sum_{j=1}^{G} \sigma_j^2 * n_{.j} = \frac{1}{505}[(77.64 * 241) + (78.81 * 264)] = 78.252$$

$$\sigma_B^2 = \frac{1}{n}\sum_{j=1}^{G} (\bar{x}_j - \bar{x}) n_{.j} = \frac{1}{505}[(9.234 - 8.803) * 241 + (8.409 - 8.803) * 264] = 0.1698$$

La varianza totale σ^2=78.422=78.252+0.1698