

## Pivot è bello

**Livello scolastico:** 1° biennio

<b>Abilità interessate</b>	<b>Conoscenze</b>	<b>Nuclei coinvolti</b>	<b>Collegamenti esterni</b>
Predisporre la struttura della matrice dei dati grezzi con riguardo a una rilevazione pianificata e inserire i dati rilevati anche in un foglio elettronico. Passare dai dati grezzi alle distribuzioni statistiche di frequenze ed alle corrispondenti rappresentazioni grafiche. Calcolare i principali valori medi e le misure di variabilità.	Distribuzioni delle frequenze a seconda del tipo di carattere. Frequenze assolute, relative, percentuali. Principali rappresentazioni grafiche per le distribuzioni di frequenze. Proprietà dei principali valori medi e delle misure di variabilità.	<u>Dati e previsioni</u>  <u>Laboratorio di matematica</u>  Argomentare, congetturare, dimostrare  Risolvere e porsi problemi	Vita sociale  Organizzazione di attività

### Contesto

Numeri, grafici.

Il contesto è di tipo matematico, in particolare riguarda l'uso di dati numerici e delle loro rappresentazioni grafiche.

Questa attività può essere introdotta, nella forma che qui viene proposta, in una classe del primo biennio quando gli studenti hanno acquisito abilità di base sull'utilizzo del foglio elettronico (selezionare, copiare, incollare, indirizzamenti assoluti e relativi, creazione guidata dei grafici, ...).

Dopo che gli studenti hanno imparato ad effettuare lo spoglio di tipo manuale di un numero limitato di semplici questionari, l'attività proposta diventa molto utile quando lo spoglio riguarda molti questionari. Trasferire i dati in una tabella di un foglio elettronico rende comodo e agevole la manipolazione flessibile di distribuzioni univariate tratte dall'intera base di dati o da un suo sottoinsieme.

### Descrizione dell'attività

Per lo sviluppo di questa attività si è utilizzato il foglio elettronico Excel.

Nell'esempio si presenta una parte di una tabella, tratta dall'unità "Arrivare a scuola" (Tabella 1) e la si descrive nelle sue parti principali al fine di chiarire i diversi concetti in essa contenuti.

Ogni riga rappresenta una unità statistica e ogni informazione da essa ricavata è disposta nelle rispettive colonne, una per la modalità di ciascun carattere.

Perché nelle colonne C e D, che contengono caratteri qualitativi sconnessi, si trovano dei numeri?

Perché proprio quei numeri?

Per poter utilizzare in modo semplice lo strumento delle tabelle pivot è importante che nelle singole celle ci siano dei valori numerici. Pertanto anche le modalità qualitative sono state codificate in forma numerica. In questo caso la scelta è stata la seguente: carattere "Sesso": modalità "Femmina" → 1, modalità "Maschio" → 2. Anche per il carattere "Comune di residenza" è stata effettuata una scelta analoga.

Tenuto conto che il questionario è stato somministrato in una scuola superiore, come è possibile trovare le età riportate in tabella? Evidentemente è stata fatta, a livello di codifica, una scelta analoga alla precedente, anche se il carattere quantitativo non lo richiedeva espressamente. E' stato infatti attribuito codice 1 all'età "Meno di 14" che corrisponde alla prima risposta data alla

domanda 1 del questionario reperibile all'unità "Arrivare a scuola", codice 2 ad una età "14", e così via. In questo modo è stata legata direttamente la modalità al numero di risposta del questionario stesso.

	A	B	C	D	E
1	codice	età	sexso	comune	classe
2	1	7	2	3	4
3	2	6	1	4	4
4	3	5	1	16	4
5	4	5	1	4	4
6	5	6	1	2	4
7	6	5	2	2	4
8	7	5	1	2	4

*Tabella 1*

Nella colonna A sono riportati i codici attribuiti ai singoli questionari in fase di spoglio. Ciò permette di individuare ogni record e di effettuare il controllo della correttezza dell'inserimento dei dati di ciascun questionario

Nella colonna B sono riportati i codici assegnati alle diverse età, secondo quanto indicato sopra: 7 corrisponde ad uno studente la cui età è di 19 anni, che è la settima risposta alla domanda 1 del questionario.

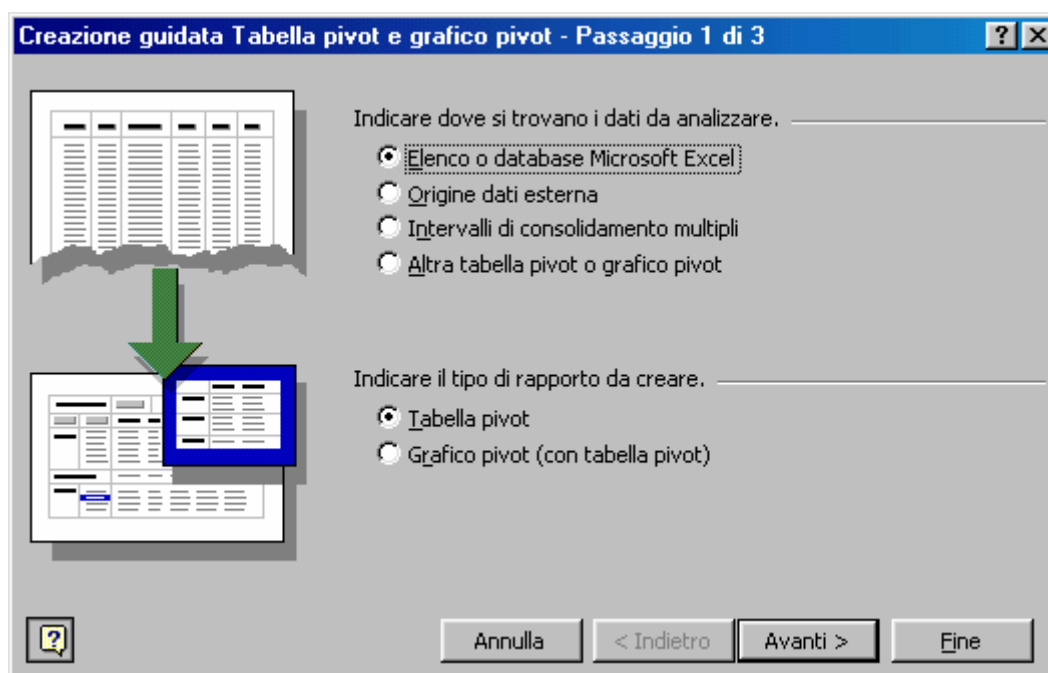
#### Prima fase

Una prima proposta riguarda il modo di ottenere informazioni circa la distribuzione univariata delle frequenze del carattere "Età" del collettivo considerato.

Si utilizza la procedura guidata per "interrogare" l'insieme dei dati oggetto di interesse (database).

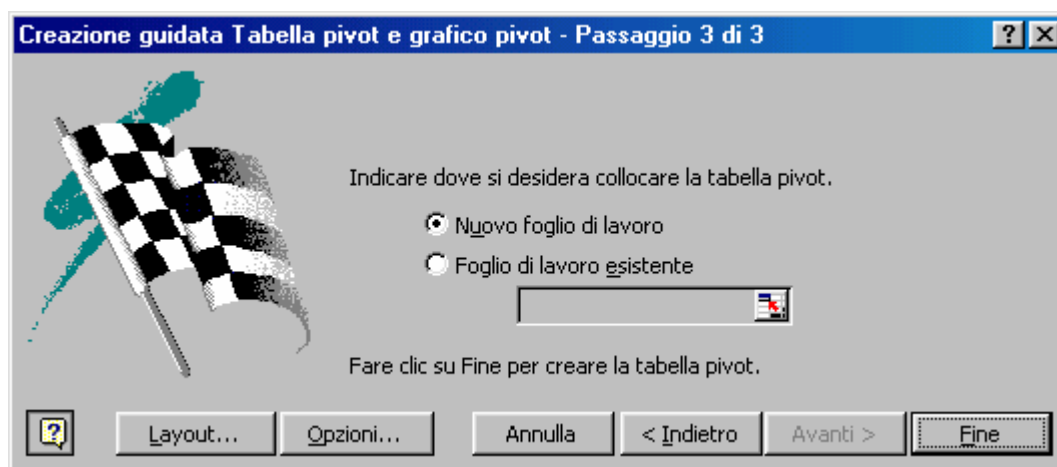
La prima cosa da fare è selezionare una cella qualsiasi del database. Successivamente si sceglie dal menu dati la voce "Report della tabella pivot e del grafico pivot ...".

Si ottiene così la seguente figura:



*Figura 1*

La scelta del pulsante “Avanti” apre una finestra che indica la zona nella quale sono residenti i dati oggetto di interesse; dopo aver confermato nuovamente con il pulsante “Avanti”, si ottiene la seguente finestra:



*Figura 2*

In questa finestra si deve scegliere dove posizionare i risultati dell’interrogazione (Tabella pivot) del database; è consigliabile, per maggior chiarezza, collocare la tabella pivot in un nuovo foglio di lavoro (come indicato in Figura 2).

Scegliendo il pulsante “Layout” si giunge alla finestra successiva che permette la costruzione della tabella pivot.



*Figura 3*

Nella finestra sono visibili i campi del database (i caratteri) e un riquadro dove, posizionando i singoli campi, è possibile ottenere la tabella desiderata. Per semplice trascinamento del campo sesso sulla zona contrassegnata dal nome Riga si inserisce il carattere “Sesso”. Trascinando il pulsante

nesso nella zona Dati si ottiene il conteggio delle unità che posseggono tali modalità. A questo riguardo è importante scegliere la modalità “Conta( )” effettuando un doppio clic sul pulsante attivo nella zona Dati dopo il trascinamento.

Per poter effettuare una interrogazione per sottoinsiemi del database, ad esempio per classe, si deve trascinare il pulsante del carattere discriminatore prescelto sul campo Pagina.

Il risultato delle scelte proposte è visibile nella figura seguente:

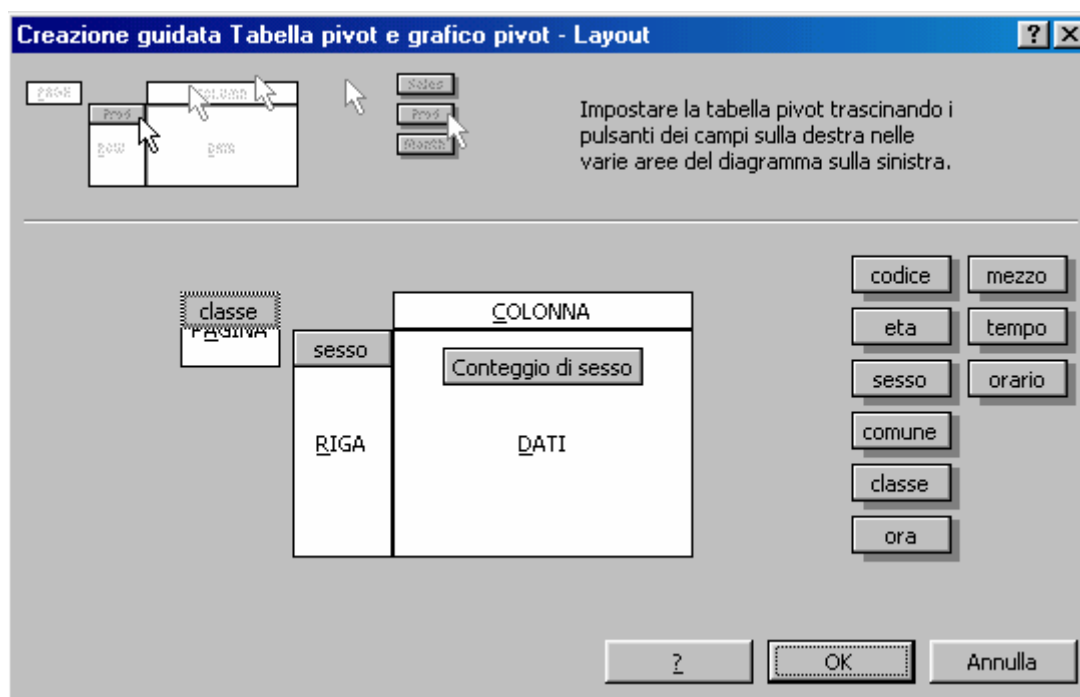


Figura 4

### Seconda fase

Il risultato della procedura illustrata nella fase precedente permette di ottenere la tabella mostrata nella figura di seguito riportata:

Microsoft Excel - arrivare a scuola			
File Modifica Visualizza Inserisci Formato Stru...			
E7 =			
	A	B	C
1	classe	(Tutto) ▼	
2			
3	Conteggio di sesso		
4	sesso ▼	Totale	
5	1	237	
6	2	109	
7	Totale complessivo	346	
8			
9			

Figura 5

La tabella pivot così ottenuta, fornisce, rispetto a tutte le classi, le informazioni richieste.

Il pulsante contenuto nella cella B1 (*nome in Excel*) permette poi di ottenere lo stesso tipo di distribuzione riferita alla modalità prescelta tra quelle proposte dal carattere “Classe”.

Bisogna tuttavia osservare che per poter lavorare sui dati ottenuti è opportuno ricopiare il contenuto della tabella e incollare i valori in un altro foglio con la procedura: “Incolla speciale”, selezionando l’opzione “Valori”. A tal fine conviene effettuare la selezione partendo da una cella in basso a destra rispetto all’area di interesse.

In questo modo si ottiene una tabella indipendente da quella pivot e soprattutto dalle modifiche determinate su di essa da ulteriori interrogazioni.

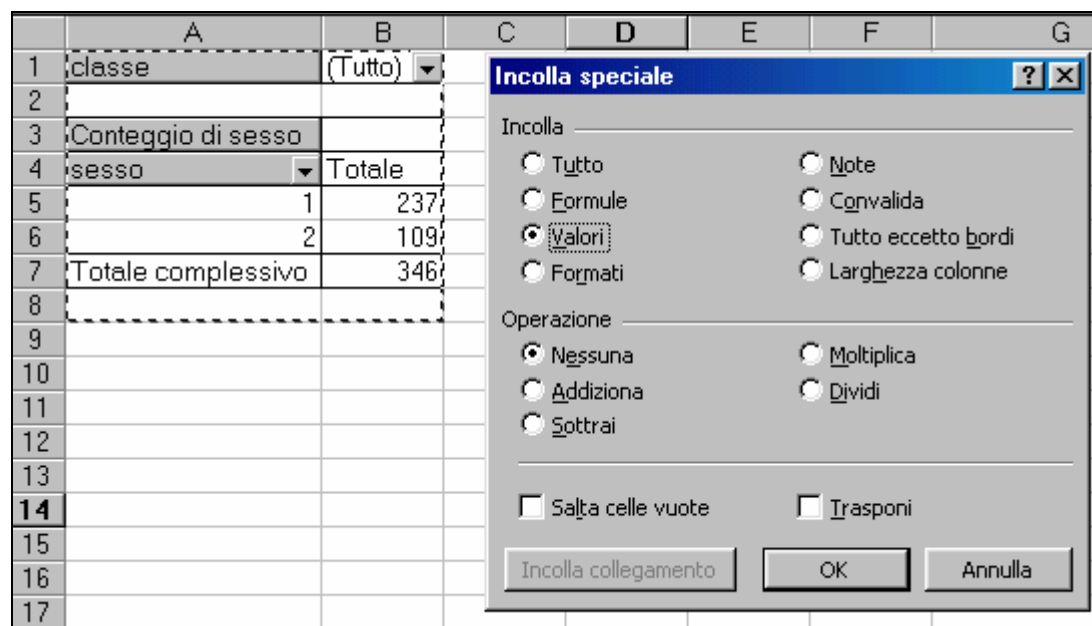


Figura 6

### Terza fase

E’ possibile rappresentare graficamente le informazioni contenute nella Figura 5? Quale tipo di grafico rappresenta in maniera più appropriata la distribuzione di frequenze esaminata?

L’insegnante sollecita la discussione e guida l’attenzione degli studenti sul significato statistico delle coppie (“Modalità del carattere Sesso”, “Frequenza”) e sulla loro rappresentazione grafica.

Una soluzione possibile può essere ottenuta mediante i seguenti passaggi.

In primo luogo occorre decodificare le modalità del carattere sesso. E’ sufficiente sovrascrivere nelle celle corrispondenti la modalità “Femmina” e la modalità “Maschio”.

Con la selezione di una cella qualsiasi all’interno della tabella pivot si può utilizzare la procedura guidata per la composizione del grafico che è in grado di rappresentare in modo corretto la distribuzione in esame.

L’insegnante, per far emergere la scelta del grafico più opportuno per rappresentare la distribuzione di frequenze, guida gli studenti a riflettere sulla natura del supporto orizzontale sul quale si trovano le modalità del sesso e sulla scelta di un grafico adeguato tra quelli proposti dal foglio elettronico per rappresentare la distribuzione considerata.

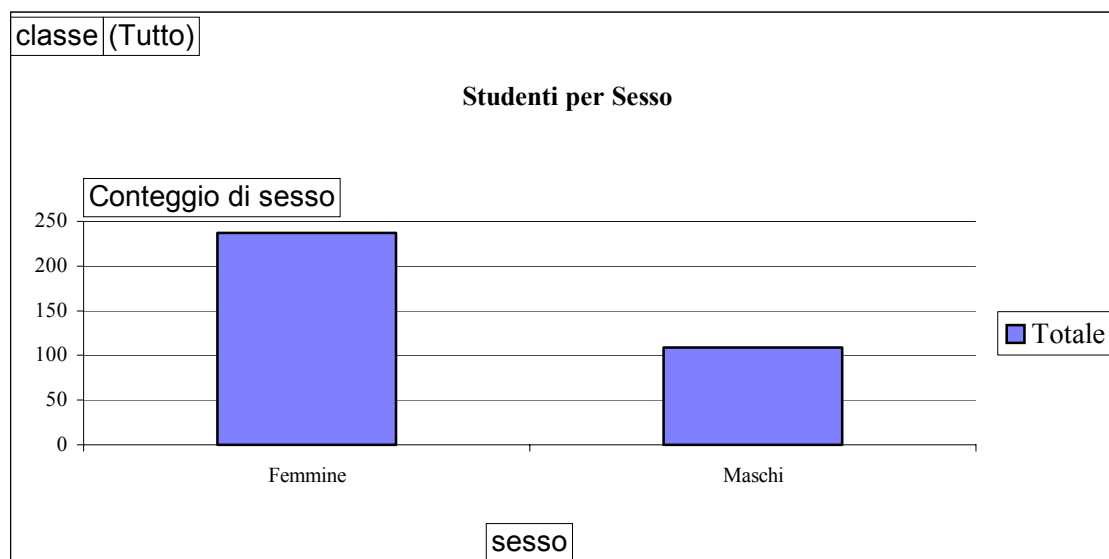


Figura 7

Che tipo di grafico propone in modo automatico il foglio elettronico? E' corretto? Perché?

#### Quarta fase

Sono di più i maschi in prima o in seconda? La tabella pivot può aiutare a rispondere a questa domanda? Si può rispondere alla prima domanda con un grafico ad aste?

L'insegnante guida gli studenti ad utilizzare la tabella pivot per trovare la risposta mediante l'interrogazione secondo la voce classe. I risultati sono riportati nelle tabelle seguenti:

classe	1
sesso	
Femmine	61
Maschi	29
Totale	90

Tabella 2

classe	2
sesso	
Femmine	36
Maschi	27
Totale	63

Tabella 3

Spontaneamente ci si attende che gli studenti rispondano che sono di più i maschi in prima. E' compito dell'insegnante guidarli al confronto fra due collettivi di diversa numerosità, pervenendo al concetto di frequenza relativa e al calcolo della Tabella 4 e della Tabella 5.

classe	1
sesso	
Femmine	67,78
Maschi	32,22
Totale	100,00

Tabella 4

classe	2
sesso	
Femmine	57,14
Maschi	42,86
Totale	100,00

Tabella 5

#### Quinta fase

Ci sono strumenti che consentono di sintetizzare una distribuzione semplice?

L'insegnante stimola l'attenzione degli studenti sull'individuazione della moda per una distribuzione di un carattere qualitativo e il calcolo della media aritmetica per un carattere quantitativo discreto.

Con l'utilizzo della tabella pivot si ottengono rapidamente le seguenti distribuzioni:

Classe	n. studenti
Prima	90
Seconda	63
Terza	117
Quarta	47
Quinta	29
Totale	346

*Tabella 6*

La Tabella 6 riporta i dati degli iscritti nelle varie classi della scuola in esame. La moda è “Terza” in quanto è la modalità che presenta la frequenza maggiore.

Se si prende in considerazione un carattere quantitativo discreto è possibile determinare la media aritmetica e ad essa associare due indici di variabilità: la varianza e lo scarto quadratico medio.

L'insegnante suggerisce agli studenti di estrarre la tabella che contiene le informazioni relative all'età degli studenti.

Età	Totale
14	61
15	69
16	107
17	61
18	40
19	5
> 19	3
Totale	346

*Tabella 7*

Qual è l'età media? La colonna dei totali contribuisce a rispondere alla domanda precedente? Come può essere trattata l'ultima modalità?

Il calcolo della media aritmetica può essere fatto anche senza l'ausilio dello strumento informatico che, tuttavia, permette di far convergere l'attenzione dello studente principalmente sugli aspetti concettuali piuttosto che su quelli calcolatori. E' fondamentale che l'insegnante induca lo studente a riflettere sul valore interpretativo degli indici trovati senza indulgere eccessivamente sulle tecniche utilizzate per la loro determinazione. A tal riguardo è importante che l'insegnante apra una discussione con gli studenti circa le soluzioni (arbitrarie) per “chiudere” l'ultima modalità.

Una soluzione possibile al problema suddetto può essere ottenuta predisponendo un'elaborazione come quella proposta nella Tabella 8.

xi	ni	xi*ni	xi - m	(xi - m) <sup>2</sup>	(xi - m) <sup>2</sup> *ni
14	61	854	-1,93	3,7385	228,0499
15	69	1035	-0,93	0,8715	60,1315
16	107	1712	0,07	0,0044	0,4728
17	61	1037	1,07	1,1374	69,3794
18	40	720	2,07	4,2703	170,8126
19	5	95	3,07	9,4033	47,0163
20	3	60	4,07	16,5362	49,6086
Totale	346	5513			625,4711
Media aritmetica (m)		15,93			
Varianza		1,81			
Scarto quad. med.		1,34			

Tabella 8

Il problema dell'ultima classe è stato risolto utilizzando informazioni aggiuntive disponibili in segreteria sull'età degli studenti di più di 19 anni.

Le prime due colonne sono state estratte dalla tabella pivot mentre le altre colonne sono state calcolate usando le formule e le funzioni di Excel.

	A	B	C	D	E	F
1	xi	ni	xi*ni	xi - m	(xi - m) <sup>2</sup>	(xi - m) <sup>2</sup> *ni
2	14	61	=A2*B2	=A2-\$C\$10	=D2*D2	=E2*B2
3	15	69	=A3*B3	=A3-\$C\$10	=D3*D3	=E3*B3
4	16	107	=A4*B4	=A4-\$C\$10	=D4*D4	=E4*B4
5	17	61	=A5*B5	=A5-\$C\$10	=D5*D5	=E5*B5
6	18	40	=A6*B6	=A6-\$C\$10	=D6*D6	=E6*B6
7	19	5	=A7*B7	=A7-\$C\$10	=D7*D7	=E7*B7
8	20	3	=A8*B8	=A8-\$C\$10	=D8*D8	=E8*B8
9	Totale	=SOMMA(B2:B8)	=SOMMA(C2:C8)			=SOMMA(F2:F8)
10	Media	aritmetica (m)	=C9/B9			
11	Varianza		=F9/B9			
12	Scarto quad. med.		=RADQ(C11)			
13						

Tabella 9

In Tabella 9 sono riportate le formule e le funzioni usate per il calcolo degli indici. L'insegnante fa notare l'uso dei riferimenti relativi e dei riferimenti assoluti nelle formule e l'impiego della funzione "Somma( )" per calcolare la somma dei valori in colonna e della funzione "Radq( )" per il calcolo della radice quadrata.

La terza colonna contiene il prodotto tra le singole modalità e le rispettive frequenze assolute. A tal riguardo l'insegnante può stimolare l'attenzione degli studenti sul fatto che tale operazione consente di "pesare" in modo diverso le singole modalità del carattere "Età".

Il rapporto tra la somma della terza colonna e la somma della seconda permette di calcolare la media aritmetica.

La quinta e la sesta colonna consentono di calcolare la varianza, ottenuta come rapporto tra la somma della sesta colonna con la somma delle seconda.



Quali informazioni si possono trarre dalla varianza? Cosa si può dire sulla sua unità di misura? E' la stessa del fenomeno del quale si vuol indagare la variabilità?

La radice quadrata della varianza fornisce lo scarto quadratico medio che è anch'esso una misura della variabilità.

Quali vantaggi ha lo scarto quadratico medio rispetto alla varianza? Quanto dista l'età 14 dalla media? E l'età 19? Possiamo esprimere queste due distanze prendendo come unità di misura lo scarto quadratico medio? Immaginiamo di avere un'altra distribuzione sempre con la stessa media 15,93 anni, e con scarto quadratico medio 3,2 anni, se uno studente ha 19 anni la sua distanza dalla media ha la stessa importanza di prima? In quale distribuzione lo studente è più vicino alla media? So lo scarto quadratico medio fosse pari a 0, cosa significherebbe?

La varianza e lo scarto quadratico medio sono misure idonee a confrontare la variabilità di due distribuzioni riferite a caratteri diversi? Perché?

### Sesta fase

C'è la possibilità di calcolare indici sintetici di una distribuzione semplice rispetto ad un carattere continuo avente modalità suddivise in classi? C'è la possibilità di calcolare la media e lo scostamento quadratico medio, analogamente a quanto si è visto nella quinta fase?

Con l'utilizzo della tabella pivot applicata sul database si ottiene rapidamente la seguente distribuzione che fa riferimento ai tempi che gli studenti impiegano per arrivare a scuola.

Tempo di percorrenza (min.)	N° studenti
fino a 15 min.	145
da 16 a 30 min.	103
da 31 a 45 min.	65
da 46 a 60 min.	24
oltre i 60 min.	9
Totale	346

*Tabella 10*

L'insegnante stimola l'attenzione degli studenti sul calcolo della media aritmetica e degli indici di variabilità in questa nuova situazione.

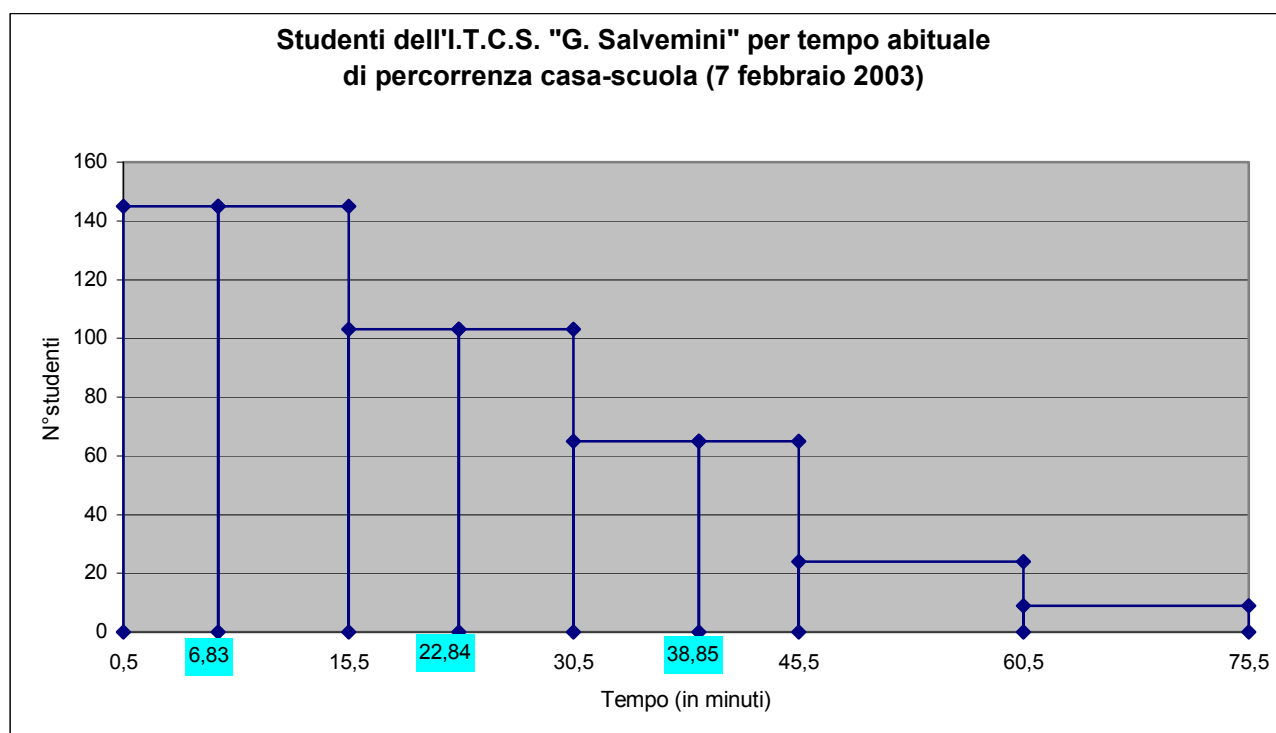
ei	es	ni	xic	xic*ni	xic - m	(xic - m) <sup>2</sup>	(xic - m) <sup>2</sup> *ni
0,5	15,5	145	8	1160	-14,84	220,08	31912,31669
15,5	30,5	103	23	2369	0,16	0,03	2,79534064
30,5	45,5	65	38	2470	15,16	229,97	14948,00682
45,5	60,5	24	53	1272	30,16	909,91	21837,87678
60,5	70	9	70	630	47,16	2224,51	20020,6142
Totale		346		7901			88721,60983
media aritmetica (m)		22,84		legenda:	ei	estremo inferiore	
varianza		256,42			es	estremo superiore	
scarto quad. medio		16,01			xic	valore centrale	

*Tabella 11*

Il caso in esame pone allo studente problemi aggiuntivi rispetto al caso precedente? Dove deve porre l'attenzione lo studente? Nell'algoritmo di calcolo o altrove?

L'insegnante guida lo studente a riflettere sulla necessità di individuare con attenzione gli estremi della prima classe e la modalità di "chiusura" dell'ultima. Nell'esempio di Tabella 5 il valore rappresentativo dell'ultima classe è stato individuato in base alle informazioni aggiuntive a disposizione (i 9 studenti appartenenti all'intervallo in esame, risiedendo tutti nello stesso comune, impiegano 70 minuti per raggiungere la scuola).

E' possibile costruire l'istogramma di frequenze? L'insegnante guida gli studenti a costruire la Figura 8, usando il grafico dispersione (XY) di Excel e dando una tabella di punti xy che rappresentano i vertici di ogni rettangolo. Dovendo rappresentare sul grafico la media, dove la si posizionerà? Sull'asse delle ascisse o delle ordinate? Perché?



*Figura 8*

Dovendo rappresentare lo scarto quadratico medio come si farà? L'insegnante guida gli studenti a considerare che lo scarto dalla media non è altro che una distanza dalla media. Dunque lo scarto quadratico medio, come media quadratica di distanze, dal punto di vista geometrico è un segmento, che è possibile disegnare sull'asse delle ascisse, a sinistra e a destra della media aritmetica. Dunque come si sono ottenuti i punti di ascissa 6,83 e 38,85? Si può calcolare qual è il numero di studenti che impiegano fra 6,83 e 38,85 minuti? Quale percentuale rappresentano sul totale degli studenti?