

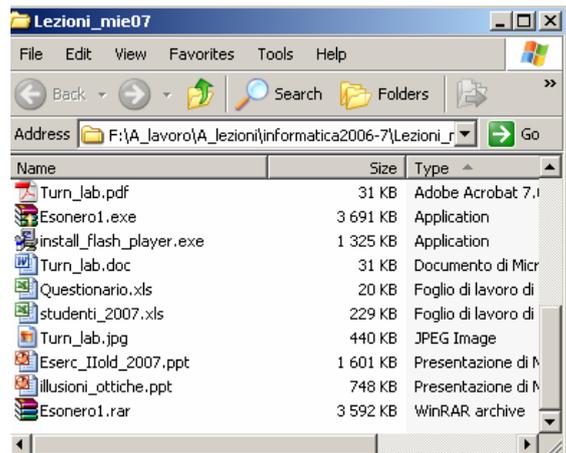
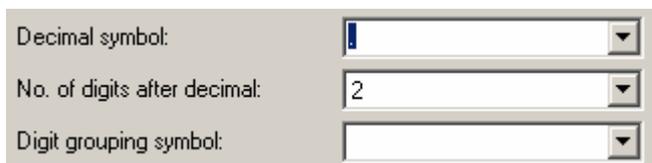
## Ulteriori conoscenze di informatica – Elementi di statistica Esercitazione I

Sui PC a disposizione sono installati diversi sistemi operativi. All'accensione scegliere Windows.

Immettere                      Nome utente    **b##**                      (## : numero del pc)  
                                          Pass.:                      **biologia##**              (## : numero del pc)  
                                                                       **Stud\_fisica**

Il sistema operativo è Windows 2000. (Non molto diverso dal WXP!) base a quanto detto a lezione controllare che:

- a) le cartelle mostrino i dettagli dei files;
- b) siano visibili le estensioni dei files anche per i tipi conosciuti;
- c) la configurazione della tastiera corrisponda alla tastiera effettivamente in uso (Inglese);
- d) i numeri decimali siano codificati con il “.” e sia usato lo spazio “ ” per il raggruppamento delle cifre.



- e) Attivare **Excel** e controllare che siano installate le opzioni “analisi dati” o aggiungerle utilizzando “strumenti > opzioni aggiuntive”
- f) Creare una cartella di lavoro nella cartella dei documenti “**UCF\_1##**” dove salvare il proprio lavoro.
- g) Scaricare e installare la cartella con i dati per l’esercitazione dalla pag. web del corso (**esercitazione1.exe**)

### Statistica descrittiva

**Esercizio 1)** Importare un file di dati in Excel, calcolo di indici di posizione, ampiezza e asimmetria. Istogramma di frequenze. Il file Excel **Esercizio1.xls** mostra una possibile soluzione e può essere utile per ricavare formule e metodi utilizzati.

Il file di dati: **dati1.txt** è un file ASCII a due colonne: in colonna 2 riporta il numero di piante della specie A osservate campionando diverse regioni indicate da un numero progressivo in colonna 1.

- Importare i dati in Excel utilizzando uno dei seguenti metodi:

- a) aprendo il file con Excel,
- b) aprendo il file con un editor di testo (blocco note) e utilizzando le opzioni copia e incolla
- c) aprendo il file da Excel utilizzando *File > apri > etc...*

Utilizzare l’opzione *dati > testo in colonne* per organizzare i dati nelle celle. **ATTENZIONE: quando si apre un file ascii con EXCEL ricordarsi di salvare i dati utilizzando il formato xls in uscita.**

- Utilizzare le funzioni di Excel per calcolare:

massimo e minimo:            **MAX(dati), MIN(dati);**  
 media, moda e mediana    **MEDIA(dati),MODA(dati),MEDIANA(dati);**  
 varianza e dev. st.            **VAR(dati), DEV.ST(dati);**  
 primo e terzo quartine      **QUARTILE(dati,Q);**  
 interquartile                 **Q3-Q1;**  
 Asimmetria e Curtosi        **ASIMMETRIA(dati),CURTOSI(dati);**

presentare i risultati in una tabella con 2 cifre significative

	A	B	C	D	E
1	N.	valore			
2	1	47		max	75.00
3	2	48		min	38.00
4	3	50			
5	4	69		media	55.16
6	5	55		moda	50.00
7	6	75		mediana	55.00
8	7	52		varianza	62.46
9	8	50		dev. st.	7.90
10	9	54			
11	10	66		Q1	50.00
12	11	55		Q3	53.75
13	12	38			
14	13	52		interquartile	9.75
15	14	49			
16	15	62		Asimmetria	0.27
17	16	44		Curtosi	0.05

Calcolare esplicitamente i momenti M2, M3 e M4 e confrontarli con i valori calcolati utilizzando le funzioni VAR, ASIMMETRIA e CURTOSI:

$(x-x)^2$	varianza (M2)	$(x-x)^3$	M3	$(x-x)^4$	M4
66.5856	61.21	-543.338	126.41	4433.642	10968.45
51.2656		-367.062		2628.162	
26.6256	dev.st.	-137.388	asimm	708.9226	curtosi
191.546	7.82	2650.991	0.26	36689.72	2.81
0.0256		-0.0041		0.000655	
393.626		7809.522		1549411	

Costruire un istogramma per visualizzare i dati: definire le classi tra minimo e massimo: 37, 41, 45,...Utilizzare quindi la funzione ISTOGRAMMA delle componenti aggiuntive “analisi dati” per calcolare l’istogramma delle frequenze:

Classe	Frequenza
37	0
41	2
45	3
49	6
53	10
57	13
61	5
65	5
69	4
73	1
77	1
Altro	0

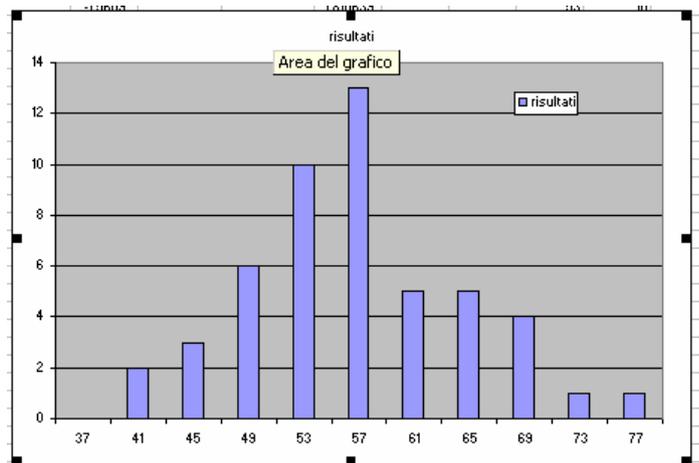
Metodo B: formule in forma di matrice

definire le classi e selezionare le celle delle frequenze, inserire la colonna dei dati e la colonna delle classi, attivare la formula in forma di matrice utilizzando **Ctrl + Shift + Enter**

The screenshot shows an Excel spreadsheet with two columns: 'classi' and 'frequenze'. The 'classi' column contains values 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77. The 'frequenze' column is empty. A formula bar shows the formula: `=FREQUENZA(matrice_dati, matrice_classi)`. Below the formula bar, the array formula is displayed: `{=FREQUENZA(B2:B51,P2:P12)}`. To the right, a smaller version of the frequency table is shown, matching the data in the screenshot.

In questo modo la tabella di frequenza viene aggiornata automaticamente se si modificano i dati o le classi. (provare)

Utilizzare le opzioni grafiche per generare l’istogramma dei dati.



**Esercizio 2)** Trattamento dati, dati multivariati e correlazione. Il file Excel **Esercizio2.xls** mostra una possibile soluzione e può essere utile per ricavare formule e metodi utilizzati.

Il file **dati2.dat** è un file ASCII a molte colonne che riporta, in funzione degli anni, le precipitazioni misurate in *pollici* in alcuni siti (6 siti) nei dintorni di Los Angeles. L'ultima colonna riporta il flusso idrico (*piedi\*acri*) nella città di Bishop(CA).

- APMAM (Mammoth Lake) inc.
- APSAB (Lake Sabrina) inc.
- APSLAKE (South Lake) inc.
- OPBPC (Big Pine Creek) inc.
- OPRC (Rock Creek) inc.
- OPSLAKE inc.
- BSAAM (Bishop(CA)) acre-feet.

I fattori conversione:  $1 \text{ inc.} = 25.4 \text{ mm}$   
 $1 \text{ acre-foot} = 1233.5 \text{ m}^3$

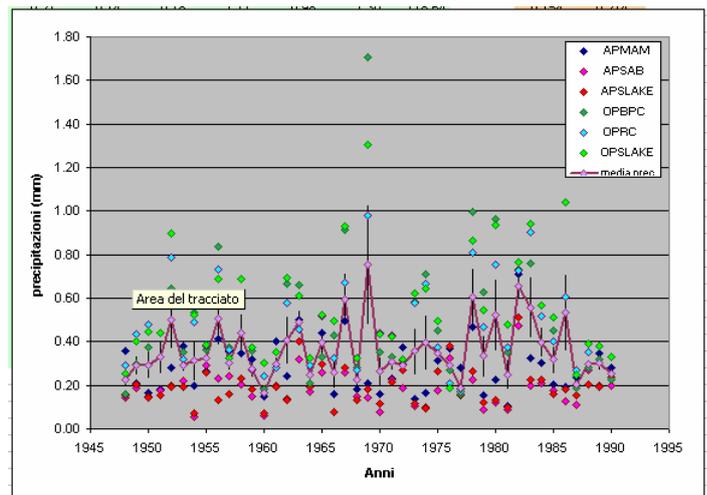
Importare i dati in Excel e convertire i dati in tabella utilizzando mm e  $\text{m}^3$  come unità di misura. Descrivere i dati con 2 o 3 cifre decimali.

	A	B	C	D	E	F	G	H	I
1	conversione		tabella MKS						
2	25.4	Year	APMAM	APSAB	APSLAKE	OPBPC	OPRC	OPSLAKE	BSAAM
3			mm	mm	mm	mm	mm	mm	$\text{m}^3$
4	1233.5	1948	0.36	0.14	0.15	0.16	0.29	0.25	43.97
5		1949	0.21	0.19	0.20	0.30	0.44	0.40	54.78
6		1950	0.17	0.15	0.14	0.37	0.48	0.45	53.64

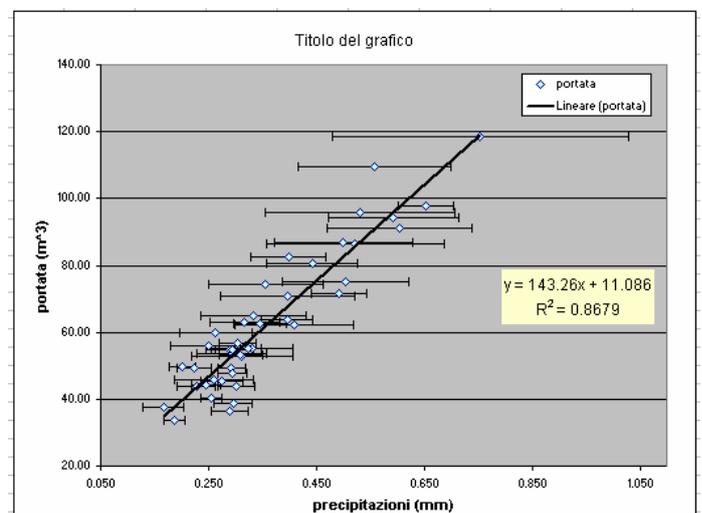
Riportare in una tabella il valor medio delle precipitazioni per ogni anno e l'errore standard sulla media ( $\sigma_x = \sigma/n^{0.5}$ ).

media prec	err. media
mm	mm
0.227	0.036
0.290	0.044
0.293	0.064

Preparare un grafico con i dati sperimentali e la media. Aggiungere una barra di errore alla media la cui ampiezza è l'errore su ciascuna delle medie.



Preparare un grafico che riporti il flusso misurato in funzione della precipitazione media. Utilizzare le opzioni grafico-linea di tendenza per riportare sul grafico la retta di regressione lineare con la formula e il parametro  $R^2$



Utilizzare gli strumenti analisi dati-correlazione per calcolare la correlazione tra le misure di precipitazioni e il flusso misurato:

	Colonna 1	Colonna 2	Colonna 3	Colonna 4	Colonna 5	Colonna 6	Colonna 7	APMAM	APSAB	APSLAKE	OPBPC	OPRC	OPSLAKE	BSAAM
Colonna 1	1							1						
Colonna 2	0.827686	1						0.827686	1					
Colonna 3	0.816076	0.900305	1					0.816076	0.900305	1				
Colonna 4	0.122386	0.039542	0.093448	1				0.122386	0.039542	0.093448	1			
Colonna 5	0.154415	0.10564	0.106384	0.864707	1			0.154415	0.10564	0.106384	0.864707	1		
Colonna 6	0.107542	0.029612	0.100587	0.943347	0.919145	1		0.107542	0.029612	0.100587	0.943347	0.919145	1	
Colonna 7	0.23857	0.183295	0.249341	0.885748	0.919627	0.938436	1	0.23857	0.183295	0.249341	0.885748	0.919627	0.938436	1

Notare che la correlazione tra flusso idrico e precipitazioni è alta nel caso di precipitazioni misurate in OPRC, OPBC e OPSLAKE e molto bassa negli altri casi. Inoltre la correlazione tra le precipitazioni misurate è alta per misure effettuate in siti vicini. (A... e O...)

Calcolare la correlazione tra precipitazioni medie e flusso idrico e osservare come questa sia eguale alla radice quadrata del parametro  $R^2$  ottenuto dalla retta di regressione lineare.

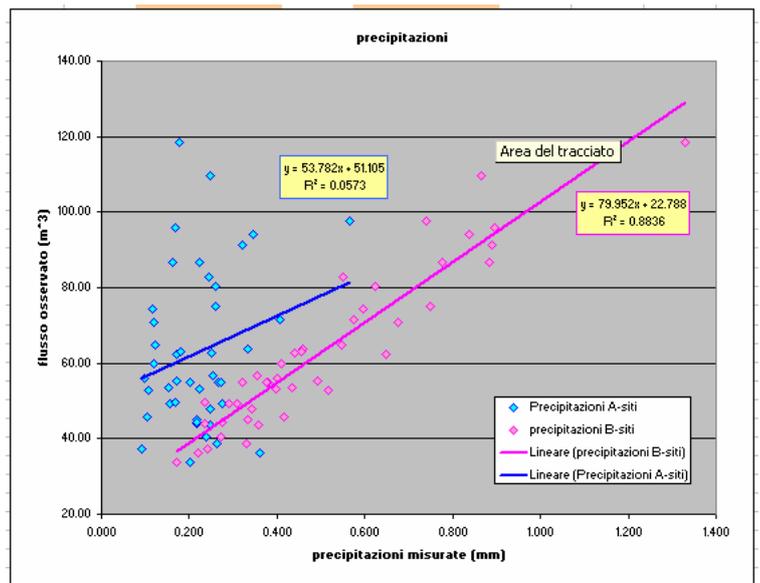
	media prec	BSAAM
media prec	1	
BSAAM	0.931613	1

correlazione media\_prec - flusso  
0.931613223

calcolare la correlazione utilizzando la funzione: CORRELAZIONE (dati\_1, dati\_2)

Riportare in due tabelle distinte le precipitazioni medie con gli errori per i siti A... e i siti O..., calcolare la correlazione con il flusso osservato e riportare su un grafico il flusso osservato in funzione delle precipitazioni medie nelle diverse regioni. Riportare le rette di regressione e i coefficienti  $R^2$

Y	Z	AA	AB	AC
A-siti		O-siti		
media prec	err. media	media prec	err. media	
mm	mm	mm	mm	
0.218	0.071	0.236	0.039	
0.201	0.006	0.380	0.042	
0.153	0.006	0.434	0.031	
0.170	0.008	0.491	0.053	
0.223	0.029	0.777	0.074	
0.246	0.050	0.357	0.020	
0.239	0.023	0.272	0.031	
correlazione		correlazione		
0.239		0.940		



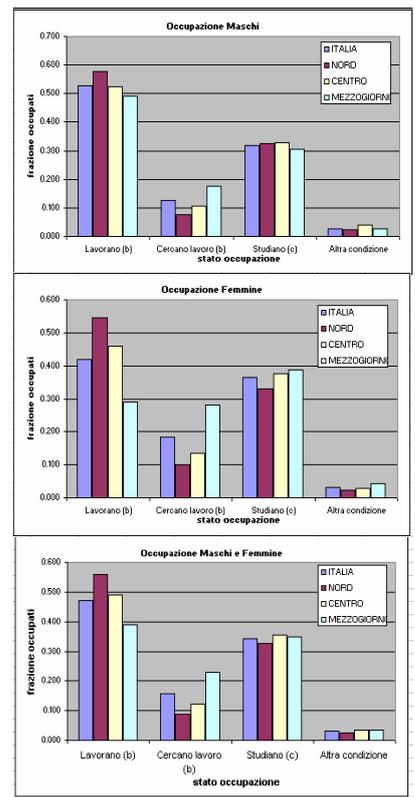
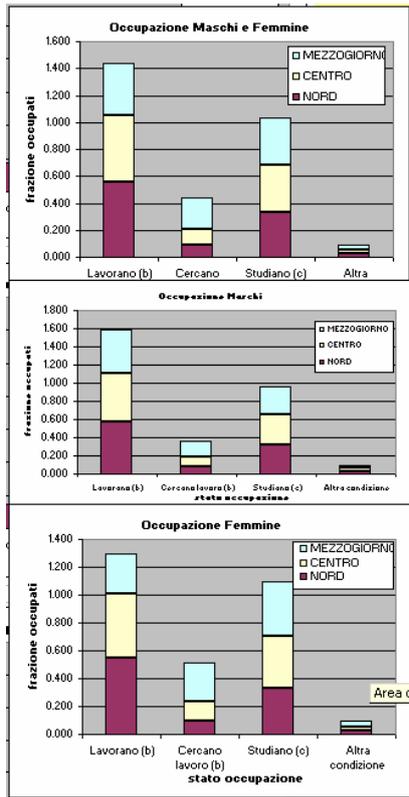
### Esercizio 3) Trattamento tabelle di contingenza

Il file **Esercizio3.xls** è un file EXCEL che contiene i dati sullo stato di occupazione di diplomati a circa tre anni dal conseguimento del diploma (Fonte: ISTAT). Il foglio **dati** contiene i dati, il foglio **esempi** contiene un esempio di soluzione, utilizzare il **foglio3** per l'esercitazione.

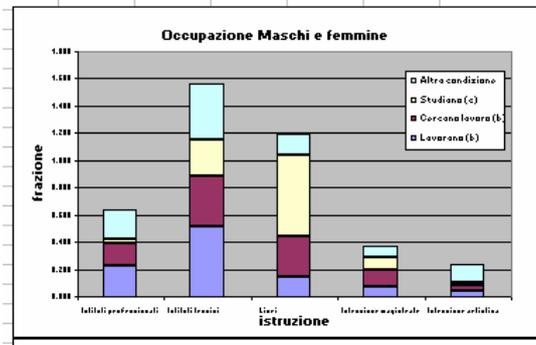


1) Per ognuna delle tabelle calcolare i profili di riga e graficare lo stato di occupazione totale (indipendentemente dall'istruzione) separatamente per maschi, femmine e per il totale della popolazione

TIPI DI SCUOLA	Valori assoluti				Composizioni percentuali %				
	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	Totale %
<b>MASCHI E FEMMINE</b>									
<b>ITALIA</b>									
Istituti professionali	48 509	11 416	5 197	2 879	0,712	0,168	0,076	0,044	100,0
Istituti tecnici	110 193	25 600	41 167	5 664	0,601	0,140	0,228	0,031	100,0
Licei	30 341	21 374	31 162	2 095	0,209	0,147	0,633	0,014	100,0
Istruzione magistrale	15 707	8 639	13 511	1 161	0,402	0,223	0,345	0,030	100,0
Istruzione artistica	7 696	3 505	2 404	1 902	0,436	0,226	0,185	0,153	100,0
<b>Totale scuole superiori</b>	<b>212 446</b>	<b>70 534</b>	<b>154 040</b>	<b>13 802</b>	<b>0,411</b>	<b>0,157</b>	<b>0,342</b>	<b>0,031</b>	<b>100,0</b>
<b>NORD</b>									
Istituti professionali	21 875	1 823	1 385	307	0,823	0,069	0,075	0,034	100,0
Istituti tecnici	46 644	4 770	16 464	1 446	0,671	0,064	0,208	0,059	100,0



2) Costruire una tabella con i profili di colonna e riportare i risultati su grafici distinti per uomini, donne e per tutta la popolazione



TIPI DI SCUOLA	Composizioni percentuali %				Totale %
	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	
<b>Maschi e Femmine</b>					
<b>ITALIA</b>					
Istituti professionali	0,228	0,162	0,034	0,216	<b>0,151</b>
Istituti tecnici	0,519	0,363	0,271	0,410	<b>0,406</b>
Licei	0,143	0,303	0,592	0,152	<b>0,322</b>
Istruzione magistrale	0,074	0,123	0,088	0,084	<b>0,087</b>
Istruzione artistica	0,036	0,050	0,016	0,138	<b>0,034</b>
<b>Totale scuole superiori</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Maschi</b>					
<b>ITALIA</b>					
Istituti professionali	0,234	0,176	0,032	0,225	<b>0,163</b>
Istituti tecnici	0,619	0,509	0,354	0,503	<b>0,518</b>
Licei	0,112	0,274	0,583	0,177	<b>0,284</b>
Istruzione magistrale	0,011	0,013	0,020	0,004	<b>0,014</b>
Istruzione artistica	0,024	0,028	0,010	0,091	<b>0,022</b>
<b>Totale scuole superiori</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Maschi e Femmine</b>					
<b>ITALIA</b>					
Istituti professionali	0,221	0,153	0,035	0,209	<b>0,140</b>
Istituti tecnici	0,401	0,270	0,204	0,338	<b>0,303</b>
Licei	0,179	0,321	0,599	0,132	<b>0,356</b>
Istruzione magistrale	0,148	0,193	0,142	0,147	<b>0,154</b>
Istruzione artistica	0,051	0,063	0,020	0,175	<b>0,046</b>
<b>Totale scuole superiori</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

3) Preparare una tabella con la distribuzione di frequenze congiunta, confrontarla con la distribuzione di frequenze che si attenderebbe nel caso di distribuzione casuale. Utilizzare, ad esempio, una tabella che mostri le differenze percentuali tra osservazioni e distribuzione random.

TIPI DI SCUOLA	Composizioni percentuali %				tabelle teoriche				differenze %						
	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	Totale %	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	Totale %	Lavorano (b)	Cercano lavoro (b)	Studiano (c)	Altra condizione	Totale %
<b>Maschi e Femmine</b>															
<b>ITALIA</b>															
Istituti professionali	0,108	0,025	0,012	0,007	<b>0,151</b>	0,071	0,024	0,052	0,005	<b>0,151</b>	34	7	-348	30	
Istituti tecnici	0,244	0,057	0,093	0,013	<b>0,406</b>	0,191	0,064	0,139	0,012	<b>0,406</b>	22	-12	-50	1	
Licei	0,067	0,047	0,202	0,005	<b>0,322</b>	0,151	0,050	0,110	0,010	<b>0,322</b>	-125	-6	46	-112	
Istruzione magistrale	0,035	0,019	0,030	0,003	<b>0,087</b>	0,041	0,014	0,030	0,003	<b>0,087</b>	-17	30	1	-3	
Istruzione artistica	0,017	0,008	0,005	0,004	<b>0,034</b>	0,017	0,005	0,012	0,001	<b>0,034</b>	5	31	-120	75	
<b>Totale scuole superiori</b>	<b>0,471</b>	<b>0,157</b>	<b>0,342</b>	<b>0,031</b>	<b>100</b>	<b>0,471</b>	<b>0,157</b>	<b>0,342</b>	<b>0,031</b>	<b>100</b>					
<b>Maschi</b>															
<b>ITALIA</b>															
Istituti professionali	0,124	0,022	0,010	0,006	<b>0,163</b>	0,086	0,021	0,052	0,005	<b>0,163</b>	31	8	-402	28	
Istituti tecnici	0,327	0,064	0,112	0,014	<b>0,518</b>	0,274	0,065	0,164	0,015	<b>0,518</b>	16	-2	-46	-3	
Licei	0,059	0,035	0,185	0,005	<b>0,284</b>	0,150	0,036	0,090	0,008	<b>0,284</b>	-154	-4	51	-60	
Istruzione magistrale	0,006	0,008	0,006	0,004	<b>0,014</b>	0,007	0,002	0,004	0,001	<b>0,014</b>	-27	-4	30	-227	
Istruzione artistica	0,012	0,004	0,003	0,003	<b>0,022</b>	0,011	0,003	0,007	0,001	<b>0,022</b>	8	22	-116	76	
<b>Totale scuole superiori</b>	<b>0,528</b>	<b>0,126</b>	<b>0,318</b>	<b>0,028</b>	<b>100</b>	<b>0,528</b>	<b>0,126</b>	<b>0,318</b>	<b>0,028</b>	<b>100</b>					
<b>Maschi e Femmine</b>															
<b>ITALIA</b>															
Istituti professionali	0,093	0,028	0,013	0,007	<b>0,140</b>	0,059	0,026	0,051	0,005	<b>0,140</b>	37	8	-303	38	
Istituti tecnici	0,160	0,050	0,074	0,011	<b>0,303</b>	0,127	0,056	0,110	0,010	<b>0,303</b>	24	-12	-48	-10	
Licei	0,075	0,058	0,218	0,004	<b>0,356</b>	0,149	0,068	0,130	0,012	<b>0,356</b>	-99	-11	40	-170	
Istruzione magistrale	0,052	0,036	0,052	0,005	<b>0,154</b>	0,064	0,028	0,056	0,005	<b>0,154</b>	-4	20	-8	-5	
Istruzione artistica	0,021	0,012	0,007	0,006	<b>0,046</b>	0,019	0,009	0,017	0,002	<b>0,046</b>	10	27	-130	74	
<b>Totale scuole superiori</b>	<b>0,418</b>	<b>0,185</b>	<b>0,364</b>	<b>0,033</b>	<b>100</b>	<b>0,418</b>	<b>0,185</b>	<b>0,364</b>	<b>0,033</b>	<b>100</b>					