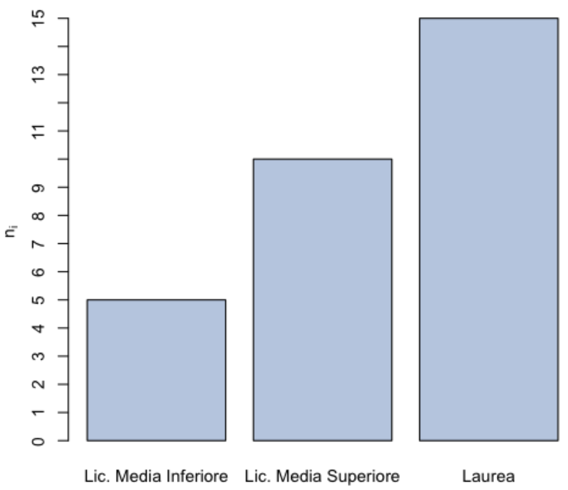
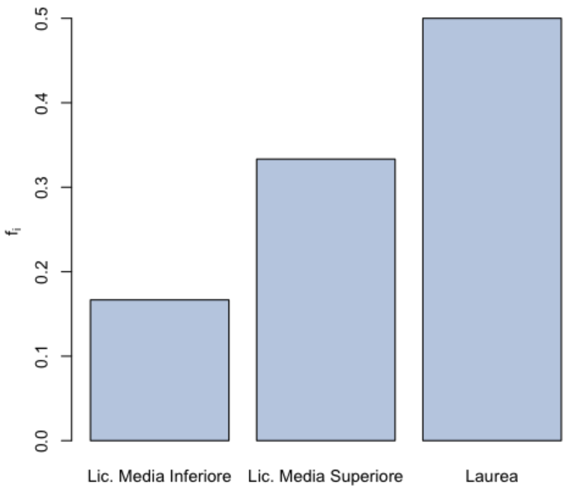
STATISTICA (I modulo - Statistica Descrittiva)
Soluzione esercitazione 2

Esercizio A.

1.

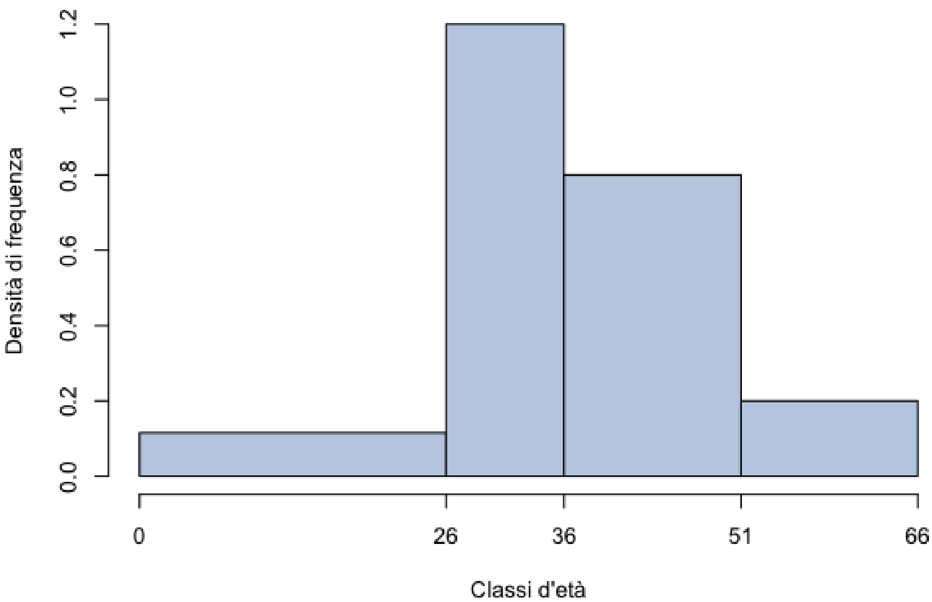
Titolo di studio	n_i	f_i
Lic. Media Inferiore	5	0,1667
Lic. Media Superiore	10	0,3333
Laurea	15	0,5000
Totale	30	1,0000





2. e 3.

Classi d'età	n_i	N_i	f_i	p_i	F_i	c_{i-1}	c_i	d_i	$h_i = n_i/d_i$
Fino a 25	3	3	0,10	10,00	0,10	0	26	26	0,1154
26-35	12	15	0,40	40,00	0,50	26	36	10	1,2000
36-50	12	27	0,40	40,00	0,90	36	51	15	0,8000
Oltre 50	3	30	0,10	10,00	1,00	51	66	15	0,2000
Totale	30		1,00	100,00					

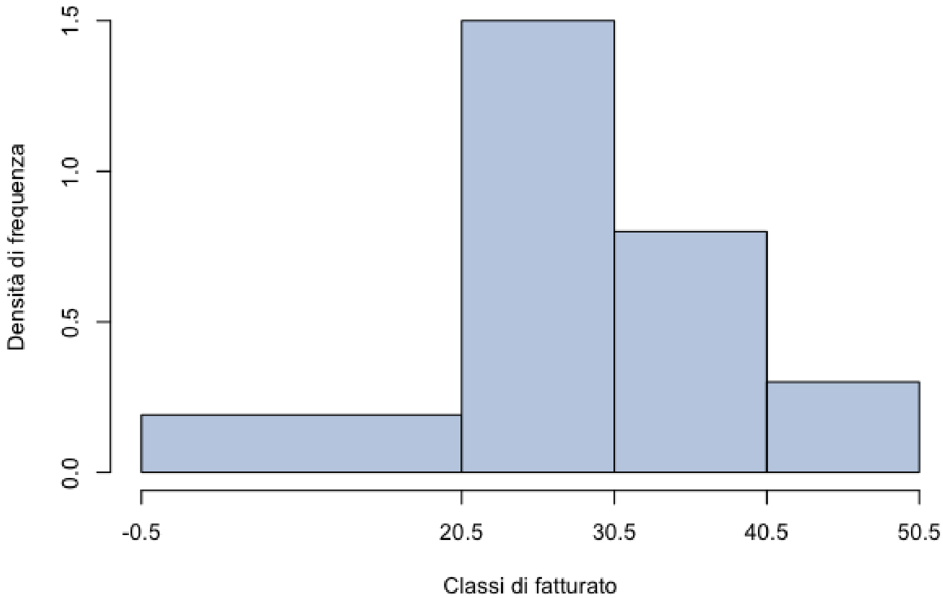


4. La frequenza dei lavoratori di età compresa tra 30 e 40 (estremi inclusi) sotto l'ipotesi di uniforme distribuzione all'interno delle classi è data da $(12 \times (36 - 30)/10 + 12 \times (41 - 36)/15) = 11{,}2$, da cui $11{,}2/30 \times 100 = 37{,}33\%$. Invece, la percentuale calcolata sulla distribuzione disaggregata fornisce il valore $8/30 \times 100 = 26{,}67\%$. Essendo i due valori piuttosto differenti possiamo dubitare in questo caso della validità dell'ipotesi di uniforme distribuzione all'interno delle classi.

5.

Classi di fatturato	n_i	N_i	f_i	p_i	F_i	c_{i-1}	c_i	d_i	$h_i = n_i/d_i$
0-20	4	4	0,1333	13,33	0,1333	-0,5	20,5	21	0,1905
21-30	15	19	0,5000	50,00	0,6333	20,5	30,5	10	1,5000
31-40	8	27	0,2667	26,67	0,9000	30,5	40,5	10	0,8000
41-50	3	30	0,1000	10,00	1,0000	40,5	50,5	10	0,3000
Totale	30		1,0000	100,00					

6.



7. Sotto l'ipotesi di uniforme distribuzione all'interno delle classi, la frequenza dei lavoratori che hanno conseguito un fatturato di almeno 25 mila € è data da $(15 \times (30{,}5 - 25)/10 + 8 + 3) = 19{,}25$, da cui $19{,}25/30 \times 100 = 64{,}17\%$. Invece, la percentuale calcolata sulla distribuzione disaggregata fornisce il valore $20/30 \times 100 = 66{,}67\%$. In questo caso l'approssimazione basata sull'ipotesi di uniforme distribuzione all'interno delle classi risulta adeguata.

8. La distribuzione di quantità (X_i) si ottiene dalla distribuzione disaggregata sommando tutti i valori del fatturato per le n_i unità che appartengono alla i-esima classe. Per esempio, la prima classe $0-20$ è costituita da $n_1 = 4$ unità statistiche, le quali hanno conseguito un fatturato pari a: $19{,}740; 19{,}472; 18{,}514; 16{,}579$. La somma di tali valori ci fornisce $X_1 = 74{,}305$. Operando in maniera analoga per tutte le classi si ottiene:

Classi di fatturato	n_i	X_i	c_{i-1}	c_i	x_i	μ_i
0-20	4	74,305	-0,5	20,5	10,0	18,576
21-30	15	377,055	20,5	30,5	25,5	25,137
31-40	8	279,476	30,5	40,5	35,5	34,934
41-50	3	141,085	40,5	50,5	45,5	47,028
Totale	30	871,921				

N.B. La seconda parte della tabella contiene valori necessari per la soluzione dell'esercizio successivo.

9. La media aritmetica del fatturato si può calcolare nei seguenti modi:

   i) a partire dalla distribuzione disaggregata

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{26{,}692 + 19{,}740 + 25{,}717 + \ldots + 28{,}670 + 48{,}428}{30} = \frac{871{,}921}{30} = 29{,}064$$

   ii) a partire dalla distribuzione di frequenza in classi ed utilizzando i valori centrali di classe $x_i = (c_{i-1} + c_i)/2$ (si veda la Tabella del punto precedente)

$$\mu = \frac{\sum_{i=1}^{k} x_i n_i}{N} = \frac{10{,}0 \times 4 + 25{,}5 \times 15 + 35{,}5 \times 8 + 45{,}5 \times 3}{30} = \frac{843}{30} = 28{,}1$$

iii) a partire dalla distribuzione di frequenza in classi ed utilizzando i valori medi di classe $\mu_i = X_i/n_i$ (si veda la Tabella del punto precedente)

$$\mu = \frac{\sum_{i=1}^{k} \mu_i n_i}{N} = \frac{18{,}57625 \times 4 + 25{,}137 \times 15 + 34{,}9345 \times 8 + 47{,}02833 \times 3}{30} = \frac{871{,}921}{30} = 29{,}064$$

Da notare che il metodo i) e iii) forniscono il medesimo risultato corretto, mentre il metodo ii) fornisce un'approssimazione che è tanto migliore quanto la distribuzione osservata si avvicina all'ipotesi di uniforme distribuzione all'interno delle classi.

Esercizio B.

1. I dati forniti costituiscono una serie di numeri indici a base fissa del tipo $I_{t|b} = a_t/a_b$ con base $b = 2004$. Per passare dai n. i. con base b alla serie dei n. i. a base fissa $b' = 2007$ occorre calcolare

$$I_{t|b'} = \frac{I_{t|b}}{I_{b'|b}} = \frac{a_t/a_b}{a_{b'}/a_b} = \frac{a_t}{a_{b'}}$$

Quindi, $I_{04|07} = I_{04|04}/I_{07|04} = 1/1{,}1259 = 0{,}88818$, $I_{05|07} = I_{05|04}/I_{07|04} = 1{,}0656/1{,}1259 = 0{,}94644$, ecc. La serie completa è la seguente:

Anno	2004	2005	2006	2007
$I_{t\|07} \times 100$	88,818	94,644	99,520	100,000

2. Un numero indice a base mobile è definito come $i_t = a_t/a_{t-1}$. A partire dalla conoscenza della serie di n. i. a base fissa si può ottenere la serie dei n. i. a base mobile come segue:

$$i_t = \frac{I_{t|b}}{I_{t-1|b}} = \frac{a_t/a_b}{a_{t-1}/a_b} = \frac{a_t}{a_{t-1}}$$

Ad esempio, il n. i. a base mobile per il 2005 è pari a $1{,}0656/1 = 1{,}0656$, quello per il 2006 è pari a $1{,}1205/1{,}0656 = 1{,}0515$, ecc.
La serie completa è la seguente:

Anno	2004	2005	2006	2007
$i_t \times 100$	-	106,56	105,15	100,48

3. La serie delle variazioni percentuali si ottiene dalla serie dei n. i. a base mobile espressi in base 1 come $v_t = (i_t - 1) \times 100$, oppure sottraendo 100 alla serie dei n. i. a base mobile espressi in base 100. Quindi:

Anno	2004	2005	2006	2007
$v_t \times 100$	-	6,56	5,15	0,48

4. Sappiamo che $a_{04} = 84088$, quindi per ricostruire la serie originaria occorre moltiplicare tale valore per la serie (espressa in base 1) dei n. i. con base 2004. Ad esempio, $a_{05} = 84088 \times 1{,}0656 = 89604$, $a_{06} = 84088 \times 1{,}1205 = 94221$, ecc. La serie completa è la seguente:

Anno	2004	2005	2006	2007
a_t	84088	89604	94221	94675

Esercizio C.

1.

Classi di età	Diploma di scuola media superiore n_i	f_i	F_i	Laurea, diploma univ., corsi post-laurea n_i	f_i	F_i
15–19	115	0,0105	0,0105	0	0,0000	0,0000
20–24	994	0,0908	0,1013	73	0,0188	0,0188
25–29	1411	0,1288	0,2301	488	0,1254	0,1442
30–34	1733	0,1582	0,3883	729	0,1874	0,3315
35–44	3372	0,3079	0,6962	1233	0,3169	0,6484
45–54	2452	0,2239	0,9201	850	0,2185	0,8669
55–64	793	0,0724	0,9925	437	0,1123	0,9792
65 e oltre	82	0,0075	1,0000	81	0,0208	1,0000
Totale	10952	1,0000		3891	1,0000	

In entrambi i casi la classe con la frequenza maggiore è la classe d'età 35–44. I lavoratori con diploma di scuola media superiore sono in generale più giovani in quanto la distribuzione delle frequenze relative cumulate F_i è sempre maggiore della corrispondente distribuzione per i lavoratori con laurea, diploma universitario o corsi post-laurea.

2. Gli individui con meno di 25 anni sono 1871 (somma delle frequenze delle celle che compaiono nelle prime due colonne della tabella), mentre tra questi quelli che possiedono al più la licenza di scuola media inferiore sono $9 + 21 + 200 + 459 = 689$. Quindi, la percentuale cercata si calcola come $689/1871 \times 100 = 36{,}825\%$.

3. La distribuzione di coloro che hanno conseguito il diploma di scuola media superiore è la seguente:

Classi d'età	c_{i-1}	c_i	x_i	n_i	f_i
15-19	15	20	17,5	115	0,0105
20-24	20	25	22,5	994	0,0908
25-29	25	30	27,5	1411	0,1288
30-34	30	35	32,5	1733	0,1582
35-44	35	45	40,0	3372	0,3079
45-54	45	55	50,0	2452	0,2239
55-64	55	65	60,0	793	0,0724
65 e oltre	65	71	68,0	82	0,0075
Totale				10952	1

dove $x_i = (c_{i-1} + c_i)/2$ sono i valori centrali di classe.

La media aritmetica è data da

$$\mu = \frac{\sum_{i=1}^{k} x_i n_i}{N} = \frac{17{,}5 \times 115 + 22{,}5 \times 994 + \ldots + 68 \times 82}{10952} = \frac{430138}{10952} = 39{,}275$$

oppure

$$\mu = \sum_{i=1}^{k} x_i f_i = 17{,}5 \times 0{,}0105 + 22{,}5 \times 0{,}0908 + \ldots + 68 \times 0{,}0075 = 39{,}275$$

Esercizio D.

1. $\frac{a_1}{b_1} + \frac{a_2}{b_2} + \ldots + \frac{a_k}{b_k} = \sum_{i=1}^{k} \frac{a_i}{b_i}$

2. $N_i = n_1 + n_2 + \ldots + n_i = \sum_{j=1}^{i} n_j$

3. $(a_1 + c) + (a_2 + c) + \ldots + (a_N + c) = \sum_{i=1}^{N} (a_i + c)$

4. $\frac{x_1 n_1 + x_2 n_2 + \ldots + x_k n_k}{n_1 + n_2 + \ldots + n_k} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i n_i}{N}$