ESERCIZIO 1

Data la seguente distribuzione percentuale delle famiglie italiane per classi di reddito, espresso in milioni di lire, (anno 1995, fonte Istat):

Classi di reddito	% famiglie
Fino a 15	5.3
15 |- 25	16.2
25 |- 35	21.1
35 |- 45	18.6
45 |- 55	13.6
Oltre 55	25.2
Totale	**100**

Assumendo 100 come valore centrale dell'ultima classe:

1) Rappresentare graficamente la distribuzione dei redditi;
2) Determinare la percentuale delle famiglie con un reddito
   a. compreso tra 20 e 25 mln di £
   b. compreso tra 15 e 20 mln di £
   c. compreso tra 15 e 17 mln di £ e tra 22 e 25 mln di £
   d. compreso tra 20 e 50 mln di £
3) Misurare la variabilità della distribuzione in termini assoluti e relativi;
4) Stabilire come si modificano media e varianza del reddito, se si ipotizza:
   a) di aumentare tutti i redditi di 10 milioni;
   b) di aumentare tutti i redditi del 10%;
   c) di convertire tutti i redditi in Euro.
5) Studiare la forma della distribuzione attraverso l'indice di Fisher

SVOLGIMENTO

1)

Per la rappresentazione grafica della distribuzione dei redditi è necessario passare alla costruzione dell'istogramma normalizzato. E' quindi necessario calcolare le densità di frequenza. Nell'ipotesi che 100 sia il valore centrale dell'ultima classe, come richiesto dall'esercizio, è agevole determinare il limite superiore della stessa:

$100 = (55 + \text{LimiteSuperiore}) / 2 \rightarrow \text{LimiteSuperiore} = (2 \times 100) - 55 = 145$

La distribuzione di partenza per la costruzione dell'istogramma normalizzato è riportata nella seguente tabella, unitamente al calcolo del valore centrale di ciascuna classe:

Classi di reddito	% famiglie	Valore Centrale Classe	Ampiezza Classe	Densità Frequenza
0 |-| 15	5.3	7.5	15	0.35
15 |- 25	16.2	20	10	1.62
25 |- 35	21.1	30	10	2.11
35 |- 45	18.6	40	10	1.86
45 |- 55	13.6	50	10	1.36
55 |-| 145	25.2	100	90	0.28
Totale	**100**			

2)

2a)

Il calcolo della percentuale di famiglie con un reddito compreso tra 20 e 25 mln di £ può essere ottenuto partendo dalla distribuzione in classi presentata al punto 1.

In particolare le famiglie di interesse sono tutte comprese nella seconda classe che va da 15 a 25 mln di £. La seconda classe ha una frequenza (percentuale) assoluta pari a 16.2 ed una densità di frequenza pari a 1.62 (frequenza/ampiezza =16.2/10).

E' quindi possibile calcolare la frequenza (percentuale) relativamente alle famiglie con reddito tra 20 e 25 mln di £ come segue:

ampiezza classe : frequenza = ampiezza sotto-intervallo : frequenza sotto-intervallo

ovvero:

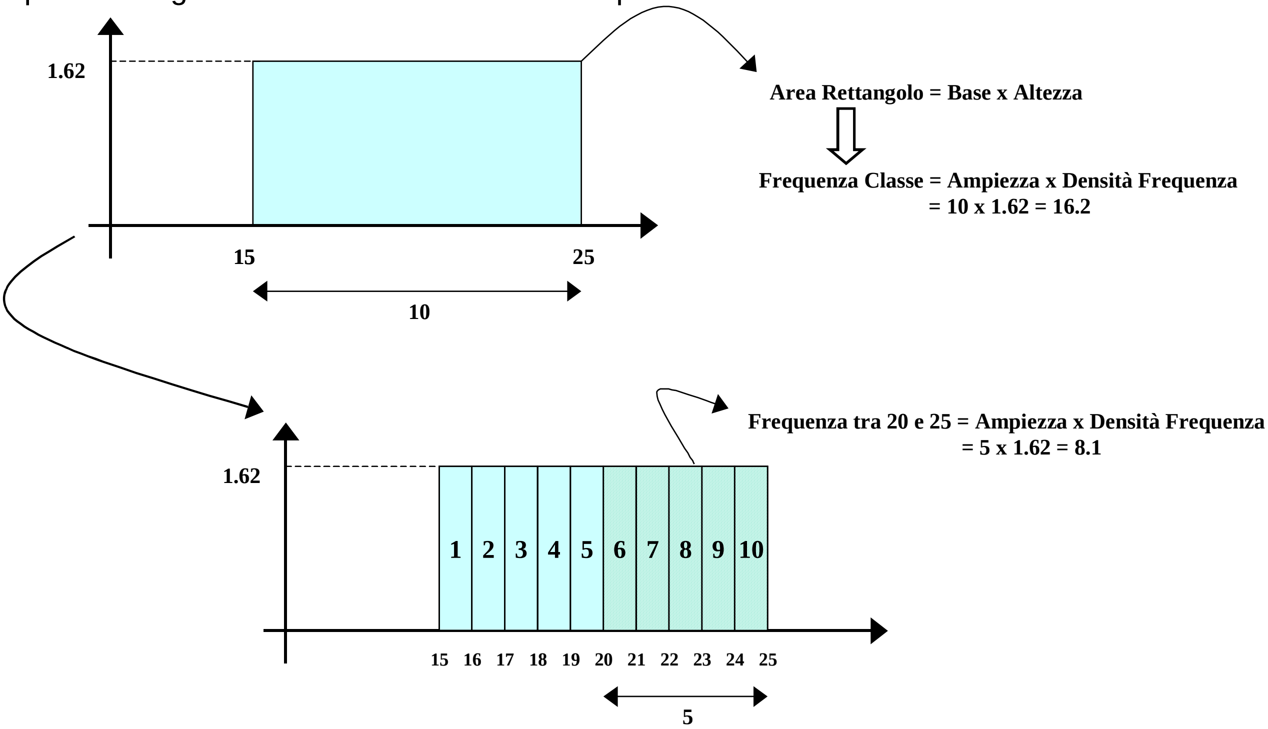
(25-15) : 16.2 = (25-20) : x

da cui si ha:

$$x = \frac{16.2 \times (25-20)}{(25-15)} = (25-20) \times \frac{16.2}{(25-15)} = (25-20) \times 1.62 = 8.1$$

Dalla precedente equazione si vede come il calcolo della frequenza relativamente alle famiglie con reddito tra 20 e 25 mln di £ può essere ottenuta direttamente moltiplicando la densità di frequenza della classe di riferimento per l'intervallo di interesse interno alla classe stessa.

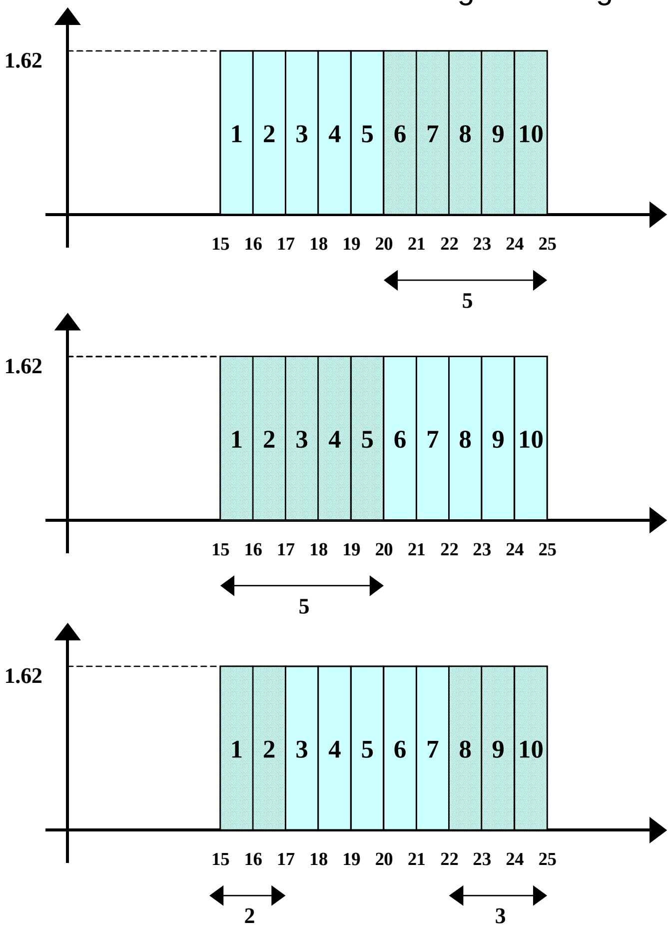
Questo è intuitivo se si tiene conto del significato di densità di frequenza, ovvero della frequenza che insiste su ciascun intervallo unitario. La seguente figura riporta il significato della densità di frequenza nel caso della classe di interesse:



Risulta quindi immediato calcolare la frequenza delle famiglie con reddito tra 20 e 25 mln di £ moltiplicando l'ampiezza dell'intervallo di interesse (pari a 5) per la densità di frequenza (pari a 1.62) ottenendo così l'area del rettangolo tratteggiato nella parte inferiore della precedente figura.

2b) e 2c)

Il calcolo di questi due punti porta allo stesso risultato di quanto ottenuto al punto 2a: ciò discende dall'ipotesi di equidistribuzione del carattere all'interno della classe.
Le famiglie con reddito tra 15 e 20 mln di £ sono infatti contenute nella seconda classe e l'intervallo di interesse è pari a 5 unità del carattere, così come del resto per le famiglie con reddito tra 15 e 17 mln di £ e tra 22 e 25 mln di £.
Quanto sopra risulta evidente dall'esame della seguente figura:



2d)

La percentuale delle famiglie con reddito compreso tra 20 e 50 mln di £ interessa invece quattro classi (dalla seconda alla quinta):

Classi di reddito	% famiglie	Valore Centrale Classe	Ampiezza Classe	Densità Frequenza
0 |-| 15	5.3	7.5	15	0.35
15 |- 25	16.2	20	10	1.62
25 |- 35	21.1	30	10	2.11
35 |- 45	18.6	40	10	1.86
45 |- 55	13.6	50	10	1.36
55 |-| 155	25.2	105	100	0.25
Totale	**100**			

Il calcolo si può scomporre facendo riferimento alle singole classi e sommando le frequenze associate a ciascuna di esse.
Per la prima classe si è già calcolata, al punto 2, la frequenza delle famiglie con reddito tra 20 e 25 mln di £, che risulta pari a 8.1
Per la terza e la quarta classe, essendo queste interamente coinvolte nel calcolo, si può considerate direttamente la frequenza percentuale associata, pari, rispettivamente a 21.1 e a 18.6.
Con riferimento alla quinta classe, la percentuale di famiglie con reddito compreso tra 45 e 50 può essere ottenuta moltiplicando l'ampiezza dell'intervallo di interesse (pari a 50-45=5) per la densità di frequenza associata alla classe (pari a 1.36), ottenendo quindi una frequenza pari a 6.8.
La frequenza della famiglie con reddito tra 20 e 50 è quindi pari a:
8.1 + 21.1 + 18.6 + 6.8 = 54.6

3)
Per il calcolo della variabilità in termini assoluti si può calcolare la varianza e lo scarto quadratico medio.
La seguente tabella riporta i dati per il calcolo della varianza sia nel caso si voglia utilizzare la formula basata sulla somma degli scarti al quadrato che quella basata sulla differenza tra momento secondo e momento primo al quadrato:

1	2	3	4	5	6	7	8	9
Classi di reddito	**Valore Centrale Classe (x_i)**	**% famiglie**	**freq. relativa (f_i)**	**x_i x f_i**	**x_i^2 x f_i**	**(x_i - μ)**	**(x_i - μ)^2**	**(x_i - μ)^2 x f_i**
0 |-| 15	7.5	5.3	0.053	0.398	2.981	-41.9075	1756.239	93.08064348
15 |- 25	20	16.2	0.162	3.240	64.800	-29.4075	864.8011	140.0977711
25 |- 35	30	21.1	0.211	6.330	189.900	-19.4075	376.6511	79.47337287
35 |- 45	40	18.6	0.186	7.440	297.600	-9.4075	88.50106	16.46119646
45 |- 55	50	13.6	0.136	6.800	340.000	0.5925	0.351056	0.04774365
55 |-| 145	100	25.2	0.252	25.200	2520.000	50.5925	2559.601	645.0194662
Totale		**100**	**1**	**49.408**	**3415.281**			**974.180**

In particolare è utile passare dalle frequenze percentuali (colonna 3) a quelle relative (colonna 4).
A questo punto nella colonna 5 è riportato il calcolo della media:

$$\mu_1 = \frac{\sum_{i=1}^{n} x_i n_i}{n} = \sum_{i=1}^{n} x_i \frac{n_i}{n} = \sum_{i=1}^{n} x_i f_i = 49.408$$

La colonna 6 riporta il calcolo del momento secondo (media dei quadrati):

$$\mu_2 = \frac{\sum_{i=1}^{n} x_i^2 n_i}{n} = \sum_{i=1}^{n} x_i^2 \frac{n_i}{n} = \sum_{i=1}^{n} x_i^2 f_i = 3415.281$$

E' quindi agevole calcolare la varianza come differenza tra il momento secondo ed il momento primo al quadrato:

$$\sigma^2 = \mu_2 - (\mu_1)^2 = 3415.281 - 49.408^2 = 974.180$$

I calcoli per la formula basata sugli scarti:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \mu_1)^2 n_i}{n} = \sum_{i=1}^{n}(x_i - \mu_1)^2 \frac{n_i}{n} = \sum_{i=1}^{n}(x_i - \mu_1)^2 f_i$$

sono riportati in colonna 7, 8 e 9 della precedente tabella.
Lo scarto quadratico medio risulta invece pari a:

$$\sigma = \sqrt{\sigma^2} = \sqrt{1106.386} = 31.212$$

Come indice di variabilità relativo è possibile calcolare il coefficiente di variazione e lo scarto quadratico medio relativo:

$$CV = \frac{\sigma}{\mu} = \frac{31.212}{49.408} = 0.63$$

$$\sigma_{rel} = \frac{\sigma}{\max(\sigma)} = \frac{\sigma}{\mu\sqrt{n-1}} = \frac{0.63}{\sqrt{100-1}} = 0.06$$

4)

Il calcolo della media e varianza nel caso in cui i dati sono soggetti a trasformazioni può essere effettuato trasformando prima i dati di partenza e poi ripetendo nuovamente i calcoli oppure sfruttando le proprietà di tali indici rispetto a trasformazioni lineari effettuate sui dati.
Di seguito si riportano entrambi i procedimenti per le tre trasformazioni richieste dall'esercizio.

4a)

Nel caso in cui tutti i redditi vengono aumentati di 10 milioni la tabella seguente riporta i calcoli per la media e varianza sui dati trasformati:

Valore Centrale Classe (x_i)	% famiglie	freq. relativa (f_i)	x_i x f_i	x_i^2 x f_i
17.5	5.3	0.053	0.928	16.231
30	16.2	0.162	4.860	145.800
40	21.1	0.211	8.440	337.600
50	18.6	0.186	9.300	465.000
60	13.6	0.136	8.160	489.600
110	25.2	0.252	27.720	3049.200
	100	**1**	**59.408**	**4503.431**

$$\mu_1 = \sum_{i=1}^{n} x_i f_i = 59.408$$

$$\sigma^2 = \mu_2 - (\mu_1)^2 = 4503.431 - 59.408^2 = 974.180$$

Come del resto era logico aspettarsi tenendo conto delle proprietà di media e varianza rispetto alla traslazione dei dati:

$$y = a + x \Rightarrow \mu_y = a + \mu_x = 10 + 49.408 = 59.408$$

$$y = a + x \Rightarrow \sigma_y^2 = \sigma_x^2 = 974.180$$

4b)
Gli stessi calcoli sono riportati nella seguente tabella nel caso in cui tutti i redditi sono aumentati del 10%:

Valore Centrale Classe (x_i)	% famiglie	freq. relativa (f_i)	x_i x f_i	x_i^2 x f_i
8.25	5.3	0.053	0.437	3.607
22	16.2	0.162	3.564	78.408
33	21.1	0.211	6.963	229.779
44	18.6	0.186	8.184	360.096
55	13.6	0.136	7.480	411.400
110	25.2	0.252	27.720	3049.200
	100	**1**	**54.348**	**4132.490**

Per aumentare tutti i redditi del 10% basta moltiplicare i valori centrali originari per la costante 1.1, come risulta dai seguenti semplici calcoli:

$$y_i = x_i + x_i \times 10\% = x_i + x_i \times \frac{10}{100} = x_i + x_i \times 0.1 = x_i \times (1+0.1) = x_i \times 1.1$$

La media e la varianza, in seguito a tale trasformazione, risultano:

$$\mu_1 = \sum_{i=1}^{n} x_i f_i = 54.348$$

$$\sigma^2 = \mu_2 - (\mu_1)^2 = 4132.490 - 54.348^2 = 1178.76$$

Ricorrendo alle proprietà di media e varianza rispetto al cambiamento di origine si ottiene più velocemente lo stesso risultato:

$$y = a \times x \Rightarrow \mu_y = a \times \mu_x = 1.1 \times 49.408 = 54.348$$

$$y = a \times x \Rightarrow \sigma_y^2 = a^2 \times \sigma_x^2 = 1.1^2 \times 974.180 = 1178.76$$

4c)
Per la trasformazione dei dati in euro, tenendo conto del differente ordine di grandezza delle due valute, è conveniente esprimere i dati in migliaia di euro. La seguente tabella riporta quindi gli analoghi calcoli effettuando la trasformazione dei dati in migliaia di euro:

Valore Centrale Classe (x_i)	Valore Centrale Classe in migliaia di € (x_i)	% famiglie	freq. relativa (f_i)	x_i x f_i	x_i^2 x f_i
7.5	3.87	5.3	0.053	0.21	0.80
20	10.33	16.2	0.162	1.67	17.28
30	15.49	21.1	0.211	3.27	50.65
40	20.66	18.6	0.186	3.84	79.38
50	25.82	13.6	0.136	3.51	90.69
100	51.65	25.2	0.252	13.01	672.15
		100	**1**	**25.52**	**910.95**

$$\mu_1 = \sum_{i=1}^{n} x_i f_i = 25.52$$

$$\sigma^2 = \mu_2 - (\mu_1)^2 = 910.95 - 25.52^2 = 259.84$$

Ricorrendo alle proprietà di media e varianza rispetto al cambiamento di origine si ottiene più velocemente lo stesso risultato. Se infatti si vuole calcolare la media e la varianza dei dati espressi in migliaia di € a partire dai valori di tali indici calcolati per i dati in milioni di £ si ha:

$$y = a \times x \Rightarrow \mu_y = a \times \mu_x = \frac{1'000}{1936.27} \times 49.408 = 25.52$$

$$y = a \times x \Rightarrow \sigma_y^2 = a^2 \times \sigma_x = \left(\frac{1'000}{1936.27}\right)^2 \times 974.180 = 259.84$$

5)

Tra i vari indici di forma si riporta di seguito il calcolo dell'indice di Fisher, basato sui momenti terzi standardizzati:

$$A_F = \sum_{i=1}^{n} \frac{z_i^3 n_i}{n} = \sum_{i=1}^{n} z_i^3 f_i = \sum_{i=1}^{n} \left(\frac{x_i - \mu_1}{\sigma}\right)^3 f_i$$

I calcoli sono riportati nella seguente tabella:

Classi di reddito	Valore Centrale Classe (x_i)	freq. relativa (f_i)	$z_i = (x_i - \mu)/\sigma$	z_i^3	$z_i^3 \times f_i$
0 |-| 15	7.5	0.053	-1.343	-2.421	-0.128
15 |- 25	20	0.162	-0.942	-0.836	-0.135
25 |- 35	30	0.211	-0.622	-0.240	-0.051
35 |- 45	40	0.186	-0.301	-0.027	-0.005
45 |- 55	50	0.136	0.019	0.000	0.000
55 |-| 145	100	0.252	1.621	4.259	1.073
Totale		**1**			0.754

Da cui risulta come la distribuzione sia caratterizzata da un'asimmetria positiva (indice maggiore di zero).
Tale risultato è confermato dal calcolo della mediana e dal confronto tra questa e la media aritmetica.
La seguente tabella riporta il calcolo delle frequenze relative cumulate (funzione di ripartizione empirica):

Classi di reddito	freq. relativa (f_i)	freq. cumulata (F_i)
0 |-| 15	0.053	0.053
15 |- 25	0.162	0.215
25 |- 35	0.211	0.426
35 |- 45	0.186	0.612
45 |- 55	0.136	0.748
55 |-| 145	0.252	1
Totale	**1**	

Da essa risulta come la mediana sia compresa nella classe di reddito tra 35 e 45 (la mediana corrisponde infatti a quel valore x per cui la funzione di ripartizione empirica F(x)=0.5) e quindi minore della media (pari a 50.668), confermando la presenza di un'asimmetria positiva.
Ciò è dovuto anche alla particolare scelta che si è fatta con riferimento al valore centrale dell'ultima classe, che tende a far aumentare notevolmente la media.
Al fine di determinare un valore approssimato per la mediana è possibile ricorrere al seguente metodo di interpolazione lineare.

Sia data la seguente notazione:

X_{me-1} - X_{me} → classe mediana

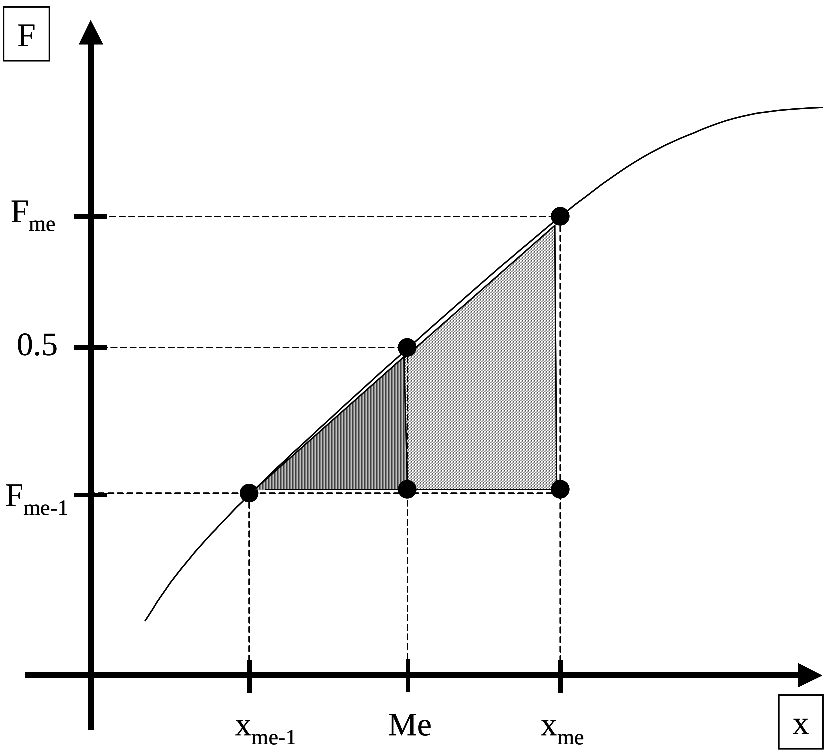
X_{me-1} → estremo inferiore della classe mediana

X_{me} → estremo superiore della classe mediana

F_{me} → valore della funzione di ripartizione empirica in corrispondenza della classe mediana

F_{me-1} → valore della funzione di ripartizione empirica in corrispondenza della classe che precede la classe mediana

Il seguente grafico riporta tali grandezze nell'ipotesi di equispaziatura dei dati all'interno della classe.



Poiché i due triangoli evidenziati in figura sono simili si ha:

$$\frac{\left(M_e - x_{me-1}\right)}{\left(0.5 - F_{me-1}\right)} = \frac{\left(x_{me} - x_{me-1}\right)}{\left(F_{me} - F_{me-1}\right)} \quad \Rightarrow \quad M_e = x_{me-1} + \frac{\left(x_{me} - x_{me-1}\right)\left(0.5 - F_{me-1}\right)}{\left(F_{me} - F_{me-1}\right)}$$

Nel caso dell'esempio in questione si ha:

$$M_e = x_{me-1} + \frac{\left(x_{me} - x_{me-1}\right)\left(0.5 - F_{me-1}\right)}{\left(F_{me} - F_{me-1}\right)} = 35 + \frac{(45-35)(0.5-0.426)}{(0.612-0.426)} = 38.98$$