Statistica I modulo (*S.I.A.F.A.*, AA 2024/25)

Le misure di sintesi

Soluzioni

1. (a) Per il calcolo della media aritmetica si considera

x_i	n_i	$x_i n_i$
98	5	490
99	4	396
100	11	1100
tot	20	1986

La media aritmetica è allora data da

$$\bar{x} = \frac{\sum_{i=1}^{3} x_i n_i}{n} = \frac{1986}{20} = 99,3$$

(b) Per la *trimmed mean* al 50% si considerano solo le 10 osservazioni centrali,

$$\bar{x}_{TR}(50\%) = \frac{99 \cdot 4 + 100 \cdot 6}{10} = 99,6$$

(c) Per la *trimmed mean* al 90% si considerano solo le 18 osservazioni centrali (in pratica si escludono le 2 osservazioni estreme)

$$\bar{x}_{TR}(90\%) = \frac{98 \cdot 4 + 99 \cdot 4 + 100 \cdot 10}{18} = 99,33$$

(d) Poiché $n = 20$ è pari, si individuano le posizioni $\frac{n}{2} = 10$ e $\frac{n}{2} + 1 = 11$. Dalla frequenze cumulate N_i,

x_i	n_i	N_i
98	5	5
99	4	9
100	11	20
tot	20	

risulta

$$Me = \frac{100 + 100}{2} = 100$$

(e) La moda è $Mo = 100$ (a questo valore corrisponde la maggiore frequenza).

(f) Sostituendo il III dato (98) con il nuovo dato (84), la nuova distribuzione di frequenze risulta

x_i	n_i	$x_i n_i$
84	1	84
98	4	392
99	4	396
100	11	1100
tot	20	1972

La media aritmetica è

$$\bar{x} = \frac{\sum_{i=1}^{3} x_i n_i}{20} = \frac{1972}{20} = 98,6$$

Nessuna variazione per la *trimmed mean* la 50% e al 90%, per la mediana e per la moda.

2. Considerando i valori calcolati nella tabella seguente

classi	n_i	c_i	$c_i n_i$	f_i	F_i	a_i	h_i
0-1	44	1	22	0,333	0,333	1	44
1-2	46	1	69	0,349	0,682	1	46
2-3	21	1	52,5	0,159	0,841	1	21
3-5	18	2	72	0,136	0,977	2	9
5-10	3	5	22,5	0,023	1	5	0,60
tot	132		238				

(a) La media aritmetica è data da

$$\bar{x} \approx \frac{\sum_{i=1}^{5} c_i n_i}{n} = \frac{238}{132} = 1,803$$

(b) Per la mediana si imposta la proporzione:

$$(2-1) : (0,682 - 0,333) = (Me - 1) : (0,50 - 0,333)$$

da cui

$$Me = 1 + \frac{0,167}{0,349} = 1,478$$

(c) La classe modale è la classe 1-2 (ad essa corrisponde la maggiore densità di frequenza)
(d) Per il I quartile si imposta la proporzione:

$$(1-0) : (0,333 - 0) = (Q_1 - 0) : (0,25 - 0)$$

da cui

$$Q_1 = \frac{0,25}{0,333} = 0,751.$$

(e) Per il 30° percentile, si imposta la proporzione

$$(1-0) : (0,333 - 0) = (P_{30} - 0) : (0,30 - 1)$$

da cui

$$P_{30} = \frac{0,30}{0,333} = 0,901.$$

3. (a) Vedi Figure **??** e **??** (poiché si tratta di distribuzioni di frequenze con classi di differente ampiezza, è necessario calcolare le densità di frequenza h_i).

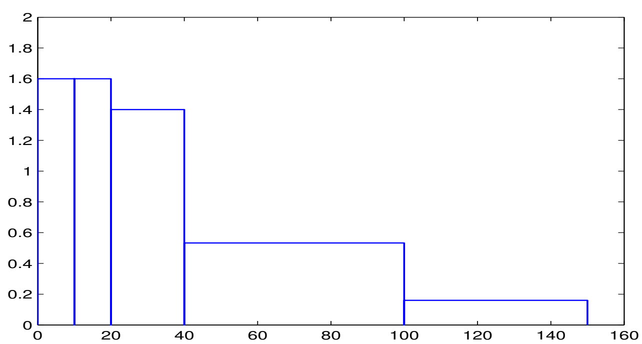

Figura 1: Istogramma famiglie italiane.

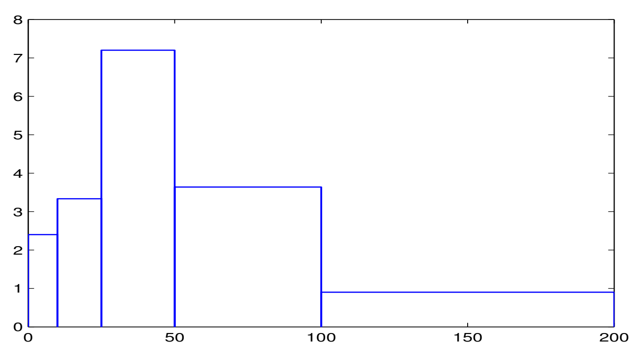

Figura 2: Istogramma famiglie statunitensi.

(b)
Per le famiglie italiane $\bar{x} = 44$.

Per le famiglie statunitensi $\bar{x} = 66,34$.

(c)
Per le famiglie italiane $Me = 32,86$.
Per le famiglie statunitensi $Me = 52,46$.

(d)
Per le famiglie italiane $P_{90} = 96,25$.
Per le famiglie statunitensi $P_{90} = 141,52$.

(e)
Per le famiglie italiane le classi modali sono 0-10 e 10-20.
Per le famiglie statunitensi la classe modale è 25-50.

4. (a) La media media per famiglia in Italia è data da

$$\frac{1400 \cdot 8000 + 1000 \cdot 6000 + 950 \cdot 5000 + 870 \cdot 2000}{8000 + 6000 + 5000 + 2000} = \frac{23690000}{21000} = 1128,095$$

5. (a) Il reddito da definire equivale al 15% percentile, ovvero P_{15}. Infatti il 15% delle famiglie ha un reddito inferiore a P_{15}. Si imposta quindi la proporzione per il calcolo del 15% percentile

$$(500 - 0) : (0,574 - 0) = (P_{15} - 0) : (0,15 - 0)$$

da cui

$$P_{15} = \frac{500 \cdot 0,15}{0,574}$$

e alla fine si ottiene

$$P_{15} = 130,662$$

(b) Il reddito da definire equivale al 90% percentile, ovvero P_{90}. Infatti il 90% delle famiglie ha un reddito inferiore a P_{90} e il restante 10% ha un reddito superiore. Si imposta quindi la proporzione per il calcolo del 90% percentile :

$$(2000 - 1000) : (0,988 - 0,845) = (P_{90} - 1000) : (0,90 - 0,845)$$

da cui

$$P_{90} - 1000 = \frac{1000 \cdot 0,055}{0,143}$$

e alla fine si ottiene

$$P_{90} = 1384,615$$

6. (a) Dopo aver individuato i coefficienti di incremento (III colonna),

Anno	TVP	Coefficienti incremento
2000	7,7	1,077
2001	7,4	1,074
2002	7,0	1,070
2003	6,9	1,069
2004	6,7	1,067
2005	6,2	1,062
2006	5,5	1,055
2007	4,0	1,040

si calcola

$$\bar{x}_g = \sqrt[8]{1,077 \cdot 1,074 \cdot \ldots \cdot 1,040} = 1,0642$$

Il tasso medio di variazione percentuale è dunque

$$\text{TMVP} = 100(1 - 1,0642) = 6,42$$

ovvero il 6,42%.