STATISTICA A-K (2014)

Soluzione esercizi da svolgere prima settimana

Classificazione di 80 aziende in base a:
X = numero di dipendenti
Y = fatturato (in milioni di euro)

X \ Y	0,5 – 1	1 – 2	2 – 4	4 – 20	Totale
10 – 20	22	3	0	0	**25**
21 – 50	2	34	4	0	**40**
51 – 100	0	0	8	2	**10**
101 – 500	0	0	0	5	**5**
Totale	**24**	**37**	**12**	**7**	**80**

- **Calcolare il numero di dipendenti medio (X) per le diverse classi di fatturato**
- **Calcolare il numero di dipendenti medio globale**
- **Rappresentare graficamente il numero di dipendenti medio in funzione delle classi di fatturato**
- **Verificare la proprietà associativa della media aritmetica**

Classificazione di 80 aziende in base a:
X = numero di dipendenti
Y = fatturato (in milioni di euro)

X \ Y	**0,5 – 1**	**1 – 2**	**2 – 4**	**4 – 20**	**Totale**
10 – 20	22	3	0	0	**25**
21 – 50	2	34	4	0	**40**
51 – 100	0	0	8	2	**10**
101 – 500	0	0	0	5	**5**
Totale	**24**	**37**	**12**	**7**	**80**

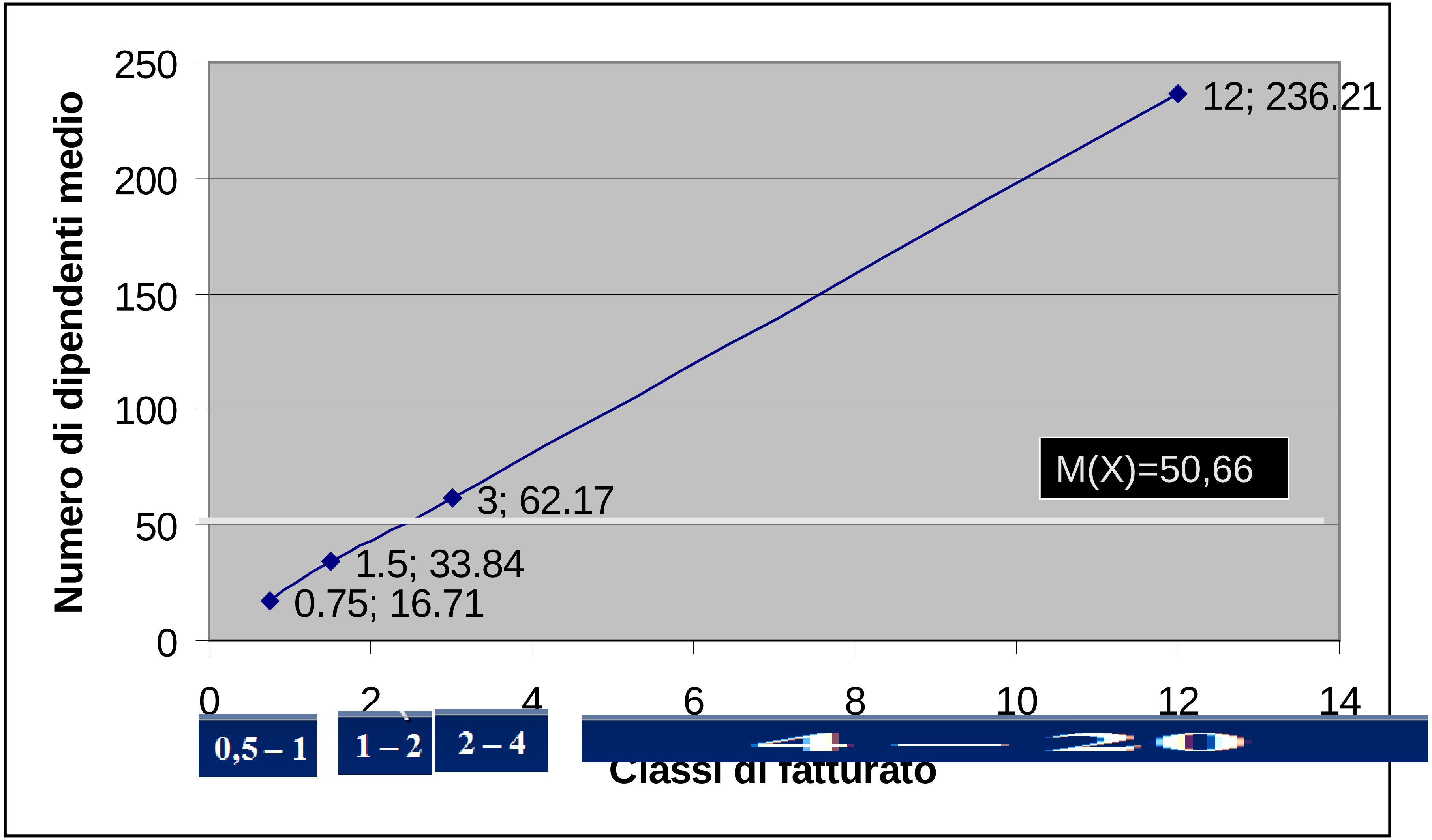
Calcolare il numero di dipendenti medio (X) per le diverse classi di fatturato

$$M(X)_1 = \frac{15 \cdot 22 + 35{,}5 \cdot 2}{24} = 16{,}71$$

$$M(X)_2 = \frac{15 \cdot 3 + 35{,}5 \cdot 34}{37} = 33{,}84$$

$M(X)_3 = 62{,}17$ $M(X)_4 = 236{,}21$

Rappresentazione grafica



Proprietà associativa (media generale dalle medie parziali)

X \ Y	0,5 – 1	1 – 2	2 – 4	4 – 20	Totale
10 – 20	22	3	0	0	**25**
21 – 50	2	34	4	0	**40**
51 – 100	0	0	8	2	**10**
101 – 500	0	0	0	5	**5**
Totale	**24**	**37**	**12**	**7**	**80**

$$M(X)_1 = \frac{15 \cdot 22 + 35{,}5 \cdot 2}{24} = 16{,}71$$

$$M(X)_2 = \frac{15 \cdot 3 + 35{,}5 \cdot 34}{37} = 33{,}84$$

$M(X)_3 = 62{,}17$ $M(X)_4 = 236{,}21$

(16,71×24+ 33,84×37+

62,17×12+
236,21×7)/80

=50,66

Esercizio (punti 6)

Nella seguente distribuzione di frequenze è riportato il numero di dipendenti di 50 aziende manifatturiere operanti in una determinata provincia.

Numero di dipendenti	Frequenze assolute
105	12
108	11
112	11
114	8
115	7
645	1

1. Si calcoli la media aritmetica, quadratica e la media geometrica del numero dei dipendenti e si commentino i risultati ottenuti.
2. Si calcoli il valore della media troncata con $\alpha=0.16$
3. Si dica quale degli indici calcolati ai punti precedenti risulta preferibile motivando la risposta.

Esercizio (punti 6)

Nella seguente distribuzione di frequenze è riportato il numero di dipendenti di 50 aziende manifatturiere operanti in una determinata provincia.

Numero di dipendenti	Frequenze assolute
105	12
108	11
112	11
114	8
115	7
645	1

1. M=120.84, M_2=142.208 M_g=114.04

$$M_g = \exp\left[\frac{1}{n}\sum_{i=1}^{r} n_i \log x_i\right] = \exp\left[\sum_{i=1}^{r} f_i \log x_i\right]. \qquad (3.19)$$

Distribuzione troncata con $\alpha=0.16$

$50\times0.16=8$ ➔ elimino i 4 valori più grandi ed i quattro valori più piccoli

Numero di dipendenti	Frequenze assolute
105	12-4=8
108	11
112	11
114	8
115	7-3=4
645	1-1=0

$M_{[0.16]}=110,2857$

In questa distribuzione c'è un chiaro valore anomalo (l'azienda con 645 dipendenti)

La media troncata è sicuramente preferibile

Es. Distribuzione di frequenze della spesa effettuata in una settimana dai titolari della carta fedeltà di un supermercato

CLASSI DI SPESA	FREQUENZA
Sino a 30	105
30-50	306
50-100	372
100-200	124
200-500	58
500-1000	32
Oltre 1000	3

Moda? Mediana? Nono decile?

Calcolo della moda

CLASSI DI SPESA	FREQUENZA	1000 × densità di frequenza
Sino a 30	105	3,5
30-50	306	15,3
50-100	372	7,44
100-200	124	1,24
200-500	58	0,19
500-1000	32	0,064
Oltre 1000	3	0
	1000	

Classe modale 30-50 Moda=40 Euro

Calcolo della mediana

CLASSI DI SPESA	FREQUENZA	f_i	F_i
Sino a 30	105	0,105	0,105
30-50	306	0,306	0,411
50-100	372	0,372	0,783
100-200	124	0,124	0,907
	1000	1	

$$x_z = \overline{x}_s + \frac{\overline{\overline{x}}_s - \overline{x}_s}{F(x_s) - F(x_{s-1})}\left[z - F(x_{s-1})\right].$$

$$x_{50\%} = 50 + \frac{100-50}{0,372}(0,5-0,411) = 61,96$$

Calcolo del nono decile

CLASSI DI SPESA	FREQUENZA	f_i	F_i
Sino a 30	105	0,105	0,105
30-50	306	0,306	0,411
50-100	372	0,372	0,783
100-200	124	0,124	0,907
200-500	58	0,058	0,965
500-1000	32	0,032	0,997
Oltre 1000	3	0,003	1
	1000	1	

$$x_{90\%} = 100 + \frac{200 - 100}{0,124}(0,9 - 0,783) = 194,35$$

Esempio (Prezzi di un bene in 4 punti vendita)

M?, Me?, DEV?, VAR?, σ?, S_M?, S_{Me}?, MAD?

x_i	$(x_i-M)^2$	$\|x_i-M\|$	$\|x_i-Me\|$
7,8	$(7{,}80-8)^2$	$\|7{,}8-8\|$	0,15
8,3	$(8{,}30-8)^2$	0,3	0,35
8,0	$(8-8)^2$	0	0,05
7,9	$(7{,}90-8)^2$	0,1	0,05
Somme	0,14	0,6	0,6

Me =*7,95*
DEV = 0,14
VAR = 0,14/4 = 0,035
σ = $(0{,}035)^{0{,}5}$
S_M = 0,6/4=0,15
S_{Me} = 0,6/4=0.15
MAD = 0,10

$$DEV = \sum_{i=1}^{n}(x_i - M)^2$$

$$VAR = \frac{\sum_{i=1}^{n}(x_i - M)^2}{n}$$

$$\sigma = \sqrt{VAR}$$

$$S_{Me} = \frac{\sum_{i=1}^{n}|x_i - Me|}{n}.$$

$$MAD = Me(|x_i - Me|).$$

$$S_M = \frac{\sum_{i=1}^{n}|x_i - M|}{n}$$

Data la seguente distribuzione

- 2 4 2 2 4 2 0 4 0 2 4 16
- Calcolare il MAD e σ sui valori originari
- Calcolare il MAD e σ sulla corrispondente distribuzione di frequenze
- Commentare i risultati ottenuti
- Quale tra i due indici risulta preferibile? Motivare la risposta

Calcolo del MAD sui valori originari $MAD=Me|x_i-Me|$

- Valori non ordinati
- 2 4 2 2 4 2 0 4 0 2 4 16
- Valori ordinati
- 0 0 2 2 2 2 2 4 4 4 4 16
- Me=2 (media tra i posti 6 e 7)
- Scostamenti dalla mediana in valore assoluto
- 2 2 0 0 0 0 0 2 2 2 2 14
- Scostamenti dalla mediana in valore assoluto ordinati
- 0 0 0 0 0 2 2 2 2 2 2 14
- MAD=2 (media tra i posti 6 e 7)

Calcolo del MAD sulla distribuzione di frequenza $MAD=Me(|x_i-Me|n_i)$

- 2 4 2 2 4 2 0 4 0 2 4 16
- Distribuzione di frequenza

x	n_i	f_i	F_i
0	2	0,167	0,167
2	5	0,417	0,584
4	4	0,333	0,917
16	1	0,083	1
	12	1	

- Me=2

Calcolo del MAD sulla distribuzione di frequenza $MAD=Me(|x_i-Me|n_i)$

x	n_i	f_i	F_i
0	2	0,167	0,167
2	5	0,417	0,584
4	4	0,333	0,917
16	1	0,083	1
	12	1	

Me=2

- Distribuzione di frequenza degli scostamenti in valore assoluto dalla mediana

$\|x_i-Me\|$	n_i	f_i	F_i
0	5	0,417	0,417
2	6	0,167+0,333	0,917
14	1	0,083	1
	12	1	

MAD=2

Osservazioni: si noti che

- Distribuzione di frequenza degli scostamenti in valore assoluto dalla mediana

$\|x_i\text{-Me}\|$	n_i	f_i	F_i
0	5	0,417	0,417
2	6	0,167+0,333	0,917
14	1	0,083	1
	12	1	

è esattamente uguale a quello che avevamo ottenuto lavorando direttamente sui dati originari

- Scostamenti dalla mediana in valore assoluto ordinati
- 0 0 0 0 0 2 2 2 2 2 2 14
- MAD=2 (media tra i posti 6 e 7)

Calcolo di σ sui valori originari

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - M)^2}{n}}$$

x_i	$(x_i\text{-}M)^2$
2	2,25
4	0,25
2	2,25
2	2,25
4	0,25
2	2,25
0	12,25
4	0,25
0	12,25
2	2,25
4	0,25
16	156,25
	193,00

- M=3,5

$\sigma = (193/12)^{0.5} = 4{,}01$

Calcolo di σ sulla distribuzione di frequenze

$$\sigma = \sqrt{\frac{\sum_{i=1}^{r}(x_i - M)^2 n_i}{n}}$$

x_i	n_i	fi	$(x_i\text{-M})^2 n_i$
0	2	0,167	24,5
2	5	0,417	11,25
4	4	0,333	1
16	1	0,083	156,25
	12	1	193

- M=3,5 σ =$(193/12)^{0.5}$=4,01

Confronto M Me

- 0 0 2 2 2 2 2 4 4 4 4 16
- Me=2 M=3,5
- Me=2 (indice di posizione robusto)
- M=3,5 (influenzata dal valore anomalo 16)

Confronto MAD σ =4,01

- 0 0 2 2 2 2 2 4 4 4 4 16
- MAD=2 σ =4,01
- MAD=2 (indice di variabilità robusto)
- σ =4,01 (influenzato dal valore anomalo 16)