Università di Cassino
Corso di Statistica 1
Esercitazione del 04/02/2008
Dott. Alfonso Piscitelli

Esercizio 1

Il seguente *data set* riporta la rilevazione di alcuni caratteri su un collettivo di 20 soggetti.

Soggetto	Età	Residenza	Reddito (Migliaia di €)	Auto Possedute	Punteggio quiz	Km giornalieri percorsi
1	22	Cantagallo	0,7	3	173	25
2	18	Cantagallo	0,2	4	168	25
3	34	Poggio a Caiano	1,6	2	165	21
4	42	Carmignano	2,5	5	180	25
5	50	Poggio a Caiano	3,2	3	163	17
6	12	Montemurlo	0,1	4	160	23
7	46	Carmignano	3,8	4	177	26
8	72	Montemurlo	1,3	2	164	35
9	27	Montemurlo	1,2	3	158	26
10	48	Carmignano	1,7	5	170	30
11	35	Montemurlo	1,9	1	167	21
12	84	Cantagallo	0,8	1	159	25
13	21	Montemurlo	0,4	5	174	26
14	44	Carmignano	1,8	4	164	33
15	56	Carmignano	1,9	2	177	24
16	58	Montemurlo	3,2	3	172	29
17	37	Cantagallo	2,1	1	166	14
18	16	Montemurlo	0,1	4	160	23
19	73	Carmignano	1,6	2	170	21
20	64	Poggio a Caiano	2,2	3	184	20

a. Determinare la differenza interquartile e lo scostamento semplice mediano per il carattere **Auto Possedute** a partire sia dalla successione di valori sia dalla distribuzione di frequenze.

b. Calcolare lo scostamento semplice medio, la varianza, lo scarto quadratico medio e il coefficiente di variazione per il carattere **Punteggio quiz**, considerando la successione di valori.

c. Determinare lo scostamento semplice medio, la varianza e lo scarto quadratico medio per il carattere **Età**, utilizzando una suddivisione in 4 classi.

Soluzioni

a) La successione dei valori ordinati in senso non decrescente e la corrispondente distribuzione di frequenza della variabile **Auto Possedute** sono le seguenti:

Soggetto	Posizione	Auto Possedute
11	1	1
12	2	1
17	3	1
3	4	2
8	5	2
15	6	2
19	7	2
1	8	3
5	9	3
9	10	3
16	11	3
20	12	3
2	13	4
6	14	4
7	15	4
14	16	4
18	17	4
4	18	5
10	19	5
13	20	5

Auto Possedute	n_i	f_i	F_i
1	3	0,15	0,15
2	4	0,2	0,35
3	5	0,25	0,60
4	5	0,25	0,85
5	3	0,15	1
Tot:	**20**	**1**	

Il primo quartile corrisponde a quel valore del carattere X che lascia alla sua sinistra il 25% delle osservazioni e alla sua destra il 75%.
Il terzo quartile corrisponde a quel valore del carattere X che lascia alla sua sinistra il 75% delle osservazioni e alla sua destra il rimanente 25%.

$$Q_1 = \frac{X_{\frac{N}{4}} + X_{\frac{N}{4}+1}}{2} = \frac{X_5 + X_6}{2} = \frac{2+2}{2} = 2$$

$$Q_3 = \frac{X_{3*\frac{N}{4}} + X_{3*\frac{N'}{4}+1}}{2} = \frac{X_{15} + X_{16}}{2} = \frac{4+4}{2} = 4$$

Nel caso delle distribuzioni di frequenza semplice, invece, i quartili vengono individuati facendo riferimento alle frequenze cumulate o alle frequenze relative cumulate. In questo caso:

. ⇒ il primo quartile è quel valore della x associato alla prima frequenza relativa cumulata maggiore di 0,25. [Q_1=2]

. ⇒ il terzo quartile è quel valore della x associato alla prima frequenza relativa cumulata maggiore di 0,75. [Q_3=4].

Si definisce differenza interquartile la differenza tra il terzo e il primo quartile.

$$IQR = Q_3 - Q_1 = 4-2=2$$

Questa quantità contiene il 50% "centrale" delle osservazioni.

Lo **scostamento semplice medio dalla Mediana** è una misura di variabilità che si ottiene come media aritmetica delle differenze, in valore assoluto, tra i valori osservati e la mediana.

$$M_e = \frac{X_{\frac{N}{2}} + X_{\frac{N}{2}+1}}{2} = \frac{X_{10}+X_{11}}{2} = \frac{3+3}{2} = 3$$

$$S_{Me} = \frac{1}{N}\sum_{i=1}^{n}|x_i - Me| = \frac{21}{20} = 1,05$$

Soggetto	Posizione	Auto Possedute	$\lvert X_i\text{-Me}\rvert$
11	1	1	2
12	2	1	2
17	3	1	2
3	4	2	1
8	5	2	1
15	6	2	1
19	7	2	1
1	8	3	0
5	9	3	0
9	10	3	0
16	11	3	0
20	12	3	0
2	13	4	1
6	14	4	1
7	15	4	1
14	16	4	1
18	17	4	1
4	18	5	2
10	19	5	2
13	20	5	2

Nel caso delle distribuzioni di frequenza semplici lo scostamento semplice dalla mediana viene calcolato con la seguente formula:

$$S_{Me} = \frac{1}{N}\sum_{i=1}^{k}|x_k - Me| \bullet n_i$$

Auto Possedute	n_i	F_i	$\lvert X_k\text{-Me}\rvert$	$\lvert X_k\text{-Me}\rvert\ n_i$
1	3	0,15	2	6
2	4	0,35	1	4
3	5	0,60	0	0
4	5	0,85	1	5
5	3	1	2	6
Tot:	**20**			**21**

b) La successione dei valori della variabile **Punteggio quiz** è la seguente:

Soggetto	Punteggio quiz	$\lvert X_i - \mu \rvert$	$(X_i-\mu)$	$(X_i-\mu)^2$
1	173	4,45	4,45	19,80
2	168	0,55	-0,55	0,30
3	165	3,55	-3,55	12,60
4	180	11,45	11,45	131,10
5	163	5,55	-5,55	30,80
6	160	8,55	-8,55	73,10
7	177	8,45	8,45	71,40
8	164	4,55	-4,55	20,70
9	158	10,55	-10,55	111,30
10	170	1,45	1,45	2,10
11	167	1,55	-1,55	2,40
12	159	9,55	-9,55	91,20
13	174	5,45	5,45	29,70
14	164	4,55	-4,55	20,70
15	177	8,45	8,45	71,40
16	172	3,45	3,45	11,90
17	166	2,55	-2,55	6,50
18	160	8,55	-8,55	73,10
19	170	1,45	1,45	2,10
20	184	15,45	15,45	238,70
	3371	**120,1**		**1020,95**

Dopo il calcolo della media della variabile **Punteggio quiz**

$$\mu = \frac{1}{N}\sum_{i=1}^{n} x_i = \frac{3371}{20} = 168,55$$

è possibile calcolare lo scostamento semplice medio.

Lo **scostamento semplice medio dalla media aritmetica** è una misura di variabilità che si ottiene come media aritmetica delle differenze, in valore assoluto, tra i valori osservati e la media aritmetica.

$$S_\mu = \frac{1}{N}\sum_{i=1}^{n} \lvert x_i - \mu \rvert = \frac{120,1}{20} = 6,005$$

La **varianza** è la media degli scarti dalla media elevati al quadrato; il rapporto tra la somma degli scarti dalla media al quadrato, e il numero delle osservazioni.

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i - \mu)^2 \qquad \sigma^2 = \frac{1020,95}{20} = 51,05$$

Una difficoltà nella interpretazione della varianza deriva dal fatto che essa è espressa nell'unità di misura del fenomeno al quadrato. Per questo motivo, si calcola lo scarto quadratico medio.

Lo **scarto quadratico medio** rappresenta la media quadratica degli scarti dalla media aritmetica.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(x_i - \mu)^2} \qquad \sigma = \sqrt{\frac{1020,95}{20}} = \sqrt{51,05} = 7,14$$

La varianza e lo scarto quadratico medio sono indici di variabilità assoluti. Un indice di variabilità relativo è il **coefficiente di variazione** definito da:

$$Cv = \frac{\sigma}{\mu}$$

$$Cv = \frac{\sigma}{\mu} = \frac{7,14}{168,55} = 0,0424$$

c) La distribuzione in classi di frequenza del carattere **Età**, è:

Età	**n$_i$**
12|-30	6
30|-48	6
48|-66	5
66|-|84	3
Tot:	**20**

Il primo passo è quello di calcolare la media aritmetica $\mu = \frac{1}{N}\sum_{i=1}^{C} x_i^c * n_i$

Età	**n$_i$**	X_i^c	(**n$_i$** * X_i^c)
12|-30	6	21	126
30|-48	6	39	234
48|-66	5	57	285
66|-|84	3	75	225
Tot:	**20**		**870**

$$\mu = \frac{870}{20} = 43,5$$

In questo caso, trattandosi di una distribuzione in classi, lo scostamento semplice medio si trova come: $S_\mu = \frac{1}{N}\sum_{i=1}^{C} \left| x_i^c - \mu \right| \bullet n_i$

Età	**n$_i$**	X_i^c	$\left\| x_i^c - \mu \right\|$	$\left\| x_i^c - \mu \right\| \bullet n_i$
12|-30	6	21	22,5	135
30|-48	6	39	4,5	27
48|-66	5	57	13,5	67,5
66|-|84	3	75	31,5	94,5
Tot:	**20**			**324**

$$S_\mu = \frac{324}{20} = 16,2$$

In questo caso, trattandosi di una distribuzione in classi, la varianza si trova come:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{C}(x_i^c - \mu)^2 * n_i$$

Età	**n_i**	X_i^c	$(x_i^c - \mu)$	$(x_i^c - \mu)^2$	$(x_i^c - \mu)^2 * n_i$
12|-30	6	21	-22,5	506,25	3038
30|-48	6	39	-4,5	20,25	121,5
48|-66	5	57	13,5	182,25	911,3
66|-|84	3	75	31,5	992,25	2977
Tot:	**20**				7047

$$\sigma^2 = \frac{7047}{20} = 352,4$$

mentre lo scarto quadratico medio è calcolabile come:

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{C}(x_i^c - \mu)^2 * n_i}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{352,4} = 18,77$$