Università di Cassino

Esercitazioni di Statistica 1 del 5 Febbraio 2010.

Dott. Mirko Bevilacqua

ESERCIZIO N° 1

A partire dalla distribuzione semplice del carattere peso rilevata su 10 studenti del corso di Microeconomia

peso: { 48, 59, 65, 66, 70, 72, 73, 75, 85, 86}

determinare media aritmetica, mediana, primo quartile, terzo quartile, campo di variazione, differenza interquartile, varianza, scarto quadratico medio, coefficiente di variazione e scarto quadratico medio relativo.

INDICI DI POSIZIONE

a) Media aritmetica

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{48+59+65+66+70+72+73+75+85+86}{10} = 69,9$$

b) Mediana

48 59 65 66 70 | 72 73 75 85 85

Essendo n pari la mediana è ottenuta come:

$$Me = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} = \frac{x_5 + x_6}{2} = \frac{70+72}{2} = 71$$

c) Primo quartile (mediana della distribuzione formata dai primi 5 valori della successione)

48 59 65 66 70 | 72 73 75 85 85

$$Q_1 = x_3 = 65$$

48 59 65 66 70

d) Terzo quartile (mediana della distribuzione formata dagli ultimi 5 valori della successione)

48 59 65 66 70 | 72 73 75 85 85

$$Q_3 = x_8 = 75$$

72 73 75 85 85

INDICI DI VARIABILITÀ

La variabilità di una distribuzione esprime la tendenza delle unità di un collettivo ad assumere diverse modalità del carattere. Per misurare la variabilità di una distribuzione è possibile utilizzare degli indici (chiamati indici di variabilità) che evidenziano aspetti particolari della variabilità di una distribuzione.

Un indice di variabilità dovrebbe soddisfare almeno due requisiti:

1) assumere il suo valore minimo se e solo se tutte le unità delle distribuzione presentano uguale modalità del carattere;

2) aumentare all'aumentare della diversità tra le modalità assunte dalle varie unità.

e) campo di variazione (indice di variabilità assoluto)

Dato un insieme di n valori osservati (X_1, X_2, ..., X_n), si definisce campo di variazione la differenza tra il più grande e il più piccolo di tali valori.

$$\textbf{Range (peso)} = X_{max} - X_{min} = 85 - 48 = 38$$

f) differenza interquartile (indice di variabilità assoluto)

Dato un insieme di n valori osservati(X_1, X_2, ..., X_n), si definisce differenza interquartile la differenza tra il primo e il terzo quartile della distribuzione.

$$IQR = Q_3 - Q_1 = 75 - 65 = 10$$

g) varianza (indice di variabilità assoluto)

*La varianza (**σ^2**) di un insieme di n valori osservati (X_1, X_2, ..., X_n) di una variabili X con media aritmetica **μ** è data da*

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

(La varianza assume valore minimo (zero) quando tutte le modalità sono uguali tra loro e aumenta all'aumentare della differenza tra i valori osservati).

σ^2 =[(48-69,9)^2+ (59-69,9)^2+(65-69,9)^2+(66-69,9)^2+(70-69,9)^2+(72-69,9)^2+(73-69,9)^2+(75-69,9)^2+(85-69,9)^2+(86-69,9)^2]/10 =116,5

P.S.: quando i dati sono organizzati in frequenza, la formula analitica per il calcolo della varianza è

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

h) scarto quadratico medio (indice di variabilità assoluto)

Lo scarto quadratico medio è dato dalla radice quadrata della varianza:

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

Con tale operazione ci si riconduce a un indice di variabilità espresso nella stessa unità di misura della variabile considerata. Come per la varianza, maggiore è la variabilità dei valori di un insieme di dati e maggiore è lo scarto quadratico medio.

(Lo scarto quadratico medio è nullo solo nel caso in cui tutti i valori sono uguali).

$$\sigma = \sqrt{\sigma^2} = \sqrt{116{,}5} = 10{,}8$$

e) coefficiente di variazione (indice di variabilità relativo)

Il coefficiente di variazione CV della distribuzione di un carattere X, di media μ>0 e scarto quadratico medio σ, è dato dal rapporto tra lo scarto quadratico medio e la media:

$$CV = \frac{\sigma}{|\mu|}$$

Tale indice non dipende dall'unità di misura e dall'ordine di grandezza della variabile. Pertanto può essere utilizzato per confrontare fenomeni espressi in unità di misura diverse o rilevati in momenti diversi. Ha il minimo uguale a zero e il massimo non definito, giacché varia al variare del tipo di distribuzione: $0 \leq CV \leq \sqrt{n-1}$.

$$CV = \frac{\sigma}{|\mu|} = \frac{10{,}8}{69{,}9} = 0{,}154$$

$$0 \leq CV \leq \sqrt{10-1} = 3$$

f) scarto quadratico medio relativo (indice di variabilità normalizzato)

Lo scarto quadratico medio relativo (σ_{REL}) si ottiene come rapporto tra il valore assunto dallo scarto ed il valore massimo che esso può assumere per la distribuzione. Quest'indice ci dice quanto siamo lontani o vicini dalla situazione di massima variabilità.

Il valore massimo che la varianza può assumere in un insieme di n unità statistiche con media pari a μ è:

$$\sigma^2_{MAX} = \mu^2 \cdot (n-1)$$

Di conseguenza, il valore massimo che può assumere lo scarto quadratico medio (in un insieme di n unità statistiche con media paria μ) è:

$$\sigma_{MAX} = \mu \cdot \sqrt{n-1}$$

$$\sigma_{REL} = \frac{\sigma_x}{MAX(\sigma_x)} = \frac{\sigma_x}{\mu_x \cdot \sqrt{n-1}}$$

$$0 \leq \sigma_{REL} \leq 1$$

$$\sigma_{REL} = \frac{\sigma_x}{\mu_x \cdot \sqrt{n-1}} = \frac{10,8}{69,9 \cdot \sqrt{10-1}} = 0,05$$

ESERCIZIO N° 2:

Si considerino le distribuzioni per classi di età della popolazione residente (espressa in migliaia) in Piemonte e in Campania nel 1979

Classi di età	**Piemonte**	**Campania**
[0 – 6]	322	603
] 6 – 14]	507	827
] 14 – 21]	477	808
] 21 – 25]	218	351
] 25 – 45]	1282	1366
] 45 – 65]	1087	1038
] 65 – 80]	703	531

Si calcolino i coefficienti di variazione per le due distribuzioni.

Quale popolazione, tra le due, ha maggiore variabilità assoluta?

Quale popolazione, tra le due, ha maggiore variabilità relativa?

PIEMONTE

Classi di età	C_i	n_i	f_i	F_i	$C_i \cdot n_i$	$C_i - \mu$	$(C_i - \mu)^2$	$(C_i - \mu)^2 \cdot n_i$
[0 – 6]	3	322	0,070	0,070	966	-35,08	1230,67	396276,50
] 6 – 14]	10	507	0,110	0,180	5070	-28,08	788,54	399789,37
] 14 – 21]	17,5	477	0,104	0,284	8347,5	-20,58	423,58	202045,32
] 21 – 25]	23	218	0,047	0,332	5014	-15,08	227,43	49580,78
] 25 – 45]	35	1282	0,279	0,611	44870	-3,08	9,49	12168,99
] 45 – 65]	55	1087	0,237	0,847	59785	16,92	286,25	311158,74
] 65 – 80]	72,5	703	0,153	1	50967,5	34,42	1184,67	832824,20
		4596 **= n**	1		**175020** $= \sum_{i=1}^{7} C_i \cdot n_i$			**2203843,89** $= \sum_{i=1}^{7} (C_i - \mu)^2 \cdot n_i$

- **Media aritmetica**

$$\mu_{PIEMONTE} = \frac{\sum_{i=1}^{7} C_i \cdot n_i}{n} = \frac{175020}{4596} = 38,08$$

oppure,

$$\mu_{PIEMONTE} = \frac{\sum_{i=1}^{7} C_i \cdot n_i}{4596} = \frac{(322 \cdot 3) + (507 \cdot 10) + ... + (703 \cdot 72,5)}{4596} = 38,08$$

- **Varianza**

$$\sigma^2_{PIEMONTE} = \frac{\sum_{i=1}^{7} (x_i - \mu)^2 n_i}{n} = \frac{2203843,89}{4596} = 479,5$$

- **Scarto quadratico medio**

$$\sigma_{PIEMONTE} = \sqrt{\sigma^2_{PIEMONTE}} = \sqrt{479,5} = 21,9$$

- **Coefficiente di variazione**

$$CV_{PIEMONTE} = \frac{\sigma_{PIEMONTE}}{|\mu_{PIEMONTE}|} = \frac{21,9}{38,08} = 0,575$$

CAMPANIA

Classi di età	C_i	n_i	f_i	F_i	$C_i \cdot n_i$	C_i-μ	$(C_i-\mu)^2$	$(C_i-\mu)^2 \cdot n_i$
[0 – 6]	3	603	0,109	0,109	1809	-28,80	829,71	500317,7
] 6 – 14]	10	827	0,150	0,259	8270	-21,80	475,45	393195,2
] 14 – 21]	17,5	808	0,146	0,405	14140	-14,30	204,63	165338,0
] 21 – 25]	23	351	0,064	0,469	8073	-8,80	77,52	27210,9
] 25 – 45]	35	1366	0,247	0,716	47810	3,20	10,21	13946,2
] 45 – 65]	55	1038	0,188	0,904	57090	23,20	538,02	558463,8
] 65 – 80]	72,5	531	0,096	1	38497,5	40,70	1656,10	879390,4
		5524 **= n**	1		**175689,5** $= \sum_{i=1}^{7} C_i \cdot n_i$			**2537862,2** $= \sum_{i=1}^{7} (C_i - \mu)^2 \cdot n_i$

- **Media aritmetica**

$$\mu_{CAMPANIA} = \frac{\sum_{i=1}^{7} C_i \cdot n_i}{n} = \frac{175689{,}5}{5524} = 31{,}8$$

- **Varianza**

$$\sigma^2_{CAMPANIA} = \frac{\sum_{i=1}^{7} (x_i - \mu)^2 n_i}{n} = \frac{253786{,}2}{5524} = 459{,}4$$

- **Scarto quadratico medio**

$$\sigma_{CAMPANIA} = \sqrt{\sigma^2_{CAMPANIA}} = \sqrt{459{,}4} = 21{,}43$$

- **Coefficiente di variazione**

$$CV_{CAMPANIA} = \frac{\sigma_{CAMPANIA}}{|\mu_{CAMPANIA}|} = \frac{21{,}43}{31{,}8} = 0{,}674$$

Come si può evincere dai valori della varianza, il Piemonte possiede una variabilità assoluta maggiore di quella della Campania. Guardando al valore del coefficiente di variazione si può concludere, diversamente da quanto si poteva evincere dalla varianza, che la Campania possiede una variabilità relativa maggiore di quella del Piemonte.

ESERCIZIO N° 3

Osserviamo nella seguente tabella la paga giornaliera in euro di 10 manovali per l'anno 2009:

Manovale	1	2	3	4	5	6	7	8	9	10
Paga giornaliera	80	65	80	55	95	60	50	65	70	70

L'amministratore della ditta edile decide che, per ogni manovale, a partire dal 1° gennaio 2010, il nuovo salario giornaliero sarà pari ad un minimo fisso di 100 euro più il 47% della retribuzione del 2009. Sapendo che la varianza per i salari del 2009 è pari a 159, determinare la varianza dei nuovi salari giornalieri.

SOLUZIONE:

Il salario giornaliero del 2010 (S_{2010}) è legato a quello del 2009 dalla relazione lineare:

$$S_{2010} = 100 + 0{,}47 \cdot S_{2009}$$

La varianza di una variabile $Y = a + bX$ può essere ottenuta a partire dalla varianza di X come:

$$\sigma^2_y = b^2\sigma^2_x$$

Quindi:

$$\sigma^2_{s_{2010}} = 0{,}47^2 \cdot 159 = 35{,}12$$

È possibile verificare il risultato ottenuto calcolando i salari giornalieri per i 10 manovali relativi all'anno 2010:

Manovale	S_{2010}	$(S_{2010}-\mu)$	$(S_{2010}-\mu)^2$
1	137,6	5,17	26,73
2	130,55	-1,88	3,53
3	137,6	5,17	26,73
4	125,85	-6,58	43,30
5	144,65	12,22	149,33
6	128,2	-4,23	17,89
7	123,5	-8,93	79,74
8	130,55	-1,88	3,53
9	132,9	0,47	0,22
10	132,9	0,47	0,22
			351,23 $= \sum_{i=1}^{n} (X_i - \mu)^2$

- **Media aritmetica**

$$\mu_{S_{2010}} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{137,6+130,55+137,6+125,85+144,65+128,2+123,5+130,55+132,9+132,9}{10} = 132,43$$

- **Varianza**

$$\sigma^2_{S_{2010}} = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n} = \frac{351,23}{10} = 35,12$$