Esercitazione 4 del corso di Statistica (parte 1)

Dott.ssa Paola Costantini

30 Ottobre 2008

Dataset dipendenti

Stipendio percepito	Età	N. di anni di servizio	Qualifica funzionale	Regime di impiego	Genere	Stato Civile	Abitazione di Proprietà
1750	34	12	Impiegato	Tempo pieno	F	Coniugato	Sì
1950	26	6	Impiegato	Tempo pieno	M	Non coniugato	No
3400	34	8	Operaio	Collaboratori esterni	F	Vedovo	Sì
2500	41	10	Operaio	Tempo pieno	F	Non coniugato	Sì
1150	38	9	Impiegato	Collaboratori esterni	M	Coniugato	No
2400	29	11	Operaio	Tempo pieno	F	Vedovo	Sì
2900	36	15	Impiegato	Tempo pieno	M	Non coniugato	Sì
2000	32	10	Impiegato	Tempo pieno	F	Vedovo	Sì
2150	36	7	Impiegato	Part time	M	Non coniugato	Sì
3900	48	8	Impiegato	Tempo pieno	F	Vedovo	Sì
1550	29	13	Dirigente	Collaboratori esterni	M	Coniugato	No
2000	31	7	Operaio	Collaboratori esterni	F	Coniugato	Sì
1800	33	8	Operaio	Tempo pieno	M	Coniugato	No
1850	42	9	Impiegato	Part time	M	Vedovo	Sì
1350	26	12	Impiegato	Collaboratori esterni	F	Coniugato	No
2450	30	15	Operaio	Tempo pieno	M	Non coniugato	Sì
2550	41	13	Impiegato	Collaboratori esterni	F	Non coniugato	Sì
2000	28	7	Impiegato	Collaboratori esterni	F	Vedovo	No
2400	33	8	Operaio	Tempo pieno	M	Non coniugato	Sì
1500	37	12	Impiegato	Collaboratori esterni	F	Coniugato	No

Esercizio 1. Si calcolino la varianza, lo scarto quadratico medio ed il coefficiente di variazione della variabile "stipendio percepito", a partire dai dati grezzi (serie di dati);

Calcolo della media e della varianza di una successione di valori:

Media

$$\boxed{\mu_X = \frac{\sum_{i=1}^{n} x_i}{n}} = 2.177,5$$

Varianza $$\boxed{\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2}$$ = 428.618,8

Scarto quadratico medio $$\boxed{\sigma = \sqrt{\sigma^2}}$$ = 654,6898

INDICI DI VARIABILITÀ RELATIVA:

Scarto quadratico medio relativo

$$\sigma_x^2 = \frac{\sigma}{\mu\sqrt{n-1}} = \frac{654{,}6898}{2177{,}8\sqrt{20-1}} = 0{,}06896$$

Si ottiene come rapporto tra il valore assunto dallo scarto ed il valore massimo che esso può assumere per la distribuzione:

$$\sigma_{rel} = \frac{\sigma_X}{\max(\sigma_X)} = \frac{\sigma_X}{\mu_X\sqrt{n-1}}$$

Coefficiente di variazione

$$CV = \frac{\sigma}{\mu} = \frac{654{,}6898}{2177{,}8} = 0{,}3006$$

E' un indice indipendente dall'unità di misura (è un numero puro) e può essere utilizzato per confrontare distribuzioni diverse.

Esercizio n 2 Si calcolino la varianza e lo scarto quadratico medio a partire dalla distribuzione del carattere "stipendio percepito" suddiviso in 4 classi equiampie.

range (STIPENDIO) = 3900 – 11500 = 2750

L'ampiezza delle classi della distribuzione di frequenza è pari a:

Ampiezza **stipendio percepito** = range/5 = 2750/5 = 550

Dati ordinati:

1150
1350
1500
1550
1750
1800
1850
1950
2000
2000
2000
2150
2400
2400
2450
2500
2550
2900
3400
3900

Le 4 classi sono, dunque, delimitate dai seguenti estremi:

C_1 = [1150; 1700]
C_2 =]1700; 2250]
C_3 =] 2250; 2800]
C_4 =]2800; 3350]
C_5 =] 3350; 3900]

x_i	ci	n_i	f_i	N_i	F_i
[1150; 1700]	1425	4	0,2	4	0,2
]1700; 2250]	1975	8	0,4	12	0,6
] 2250; 2800]	2525	5	0,25	17	0,85
]2800; 3350]	3075	1	0,05	18	0,9
] 3350; 3900]	3625	2	0,1	20	1
Totali		**20**	**1**		

$\mu = 2222{,}5$

$$\sigma_X^2 = \frac{1}{n}\sum_{i=1}^{k}\left(c_i - \mu\right)^2 n_i \quad =$$

$$\sigma_x^2 = \frac{1}{20}\sum_{x=1}^{n}\left[(1425-2222{,}5)^2 \times 4 + (1975-2222{,}5)^2 \times 8 + ...+ (3625-2222{,}5)^2 \times 2\right] = \frac{8152375}{20} = 407618{,}75$$

$$\sigma = \sqrt{407618{,}75} = 638{,}45$$

Esercizio 3. A partire dai dati ordinati, per le prime 5 modalità del carattere "Stipendio percepito" rispettivamente delle categorie *impiegati* ed *operai*:

a) costruire la curva di Lorenz;
b) determinare i rapporti di concentrazione.

a)

Primi 5 stipendi categoria Impiegati

x_i	n_i	f_i	p_i=Fi	$x_h \cdot n_h$	$S_i = \sum_{h=1}^{i} x_h \cdot n_h$	q_i
1150	1	0,2	0,2	1150	1150	0,151
1350	1	0,2	0,4	1350	2500	0,328
1500	1	0,2	0,6	1500	4000	0,526
1750	1	0,2	0,8	1750	5750	0,756
1850	1	0,2	1	1850	7600	1
	n = 5	**1**				

Si osservi che $R = \frac{\Delta}{2\mu}$, da cui:

$$\Delta = \frac{\begin{array}{l} 2\{|1150-1350| \times 1 \times 1 + |1150-1500| \times 1 \times 1 + |1150-1750| \times 1 \times 1 + |1150-1850| \times 1 \times 1 + |1350-1500| \times 1 \times 1 + \\ |1350-1750| \times 1 \times 1 + |1350-1850| \times 1 \times 1 + |1500-1750| \times 1 \times 1 + |1500-1850| \times 1 \times 1 + |1750-1850| \times 1 \times 1\} \end{array}}{5^2 - 5}$$

$$= 2 \times \frac{3600}{20} = 360$$

Altro procedimento:

Per il calcolo del Δ si può costruire la tabella delle differenze semplici, come segue:

	1150	1350	1500	1750	1850
1150	0	-200	-350	-600	-700
1350	200	0	-150	-400	-500
1500	350	150	0	-250	-350
1750	600	400	250	0	-100
1850	700	500	350	100	0

per poi applicare la formula:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{n(n-1)} = \frac{2\times(200+350+600+\cdots 250+350+100)}{5\times 4} = \frac{7200}{20} = 360$$

$\mu = 1520$

Quindi $R = \frac{360}{2\times 1520} = 0{,}12$

Tale risultato indica che nella distribuzione osservata la concentrazione è circa il 12% della massima concentrazione che avrebbe potuto rilevarsi.

Primi 5 stipendi categoria Operai

x_i	n_i	f_i	N_i	$p_i = F_i$	$x_h \cdot n_h$	$S_i = \sum_{h=1}^{i} x_h \cdot n_h$	q_i
1800	1	0,2	1	0,2	1800	1800	0,157
2400	3	0,6	4	0,8	7200	9000	0,786
2450	1	0,2	5	1	2450	11450	1
	5	**1**			**11450**		

$$\Delta = \frac{2\{\left|1800-2400\right|\times 1\times 3 + \left|1800-2450\right|\times 1\times 1 + \left|2400-2450\right|\times 3\times 1\}}{5^2-5}$$

$$= 2\times\frac{2600}{20} = 260$$

Altro procedimento:

Per il calcolo del Δ si può costruire la tabella delle differenze semplici, come segue:

		1	3	1
		1800	2400	2450
1	1800	0	1800	650
3	2400	1800	0	150
1	2450	650	150	0

per poi applicare la formula:

$$\Delta = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j| \cdot n_i \cdot n_j}{n(n-1)} = \frac{2\times(1800+650+150)}{5\times 4} = \frac{5200}{20} = 260$$

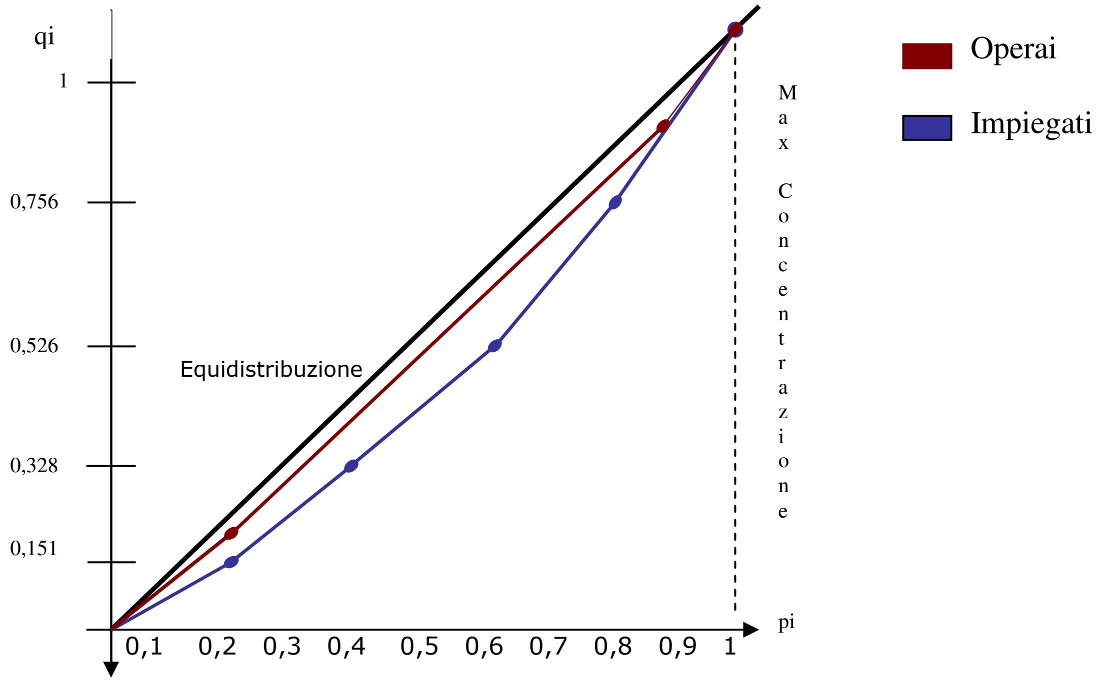
$\mu = 2290$

Quindi $R = \frac{260}{2\times 2290}$ =0,0567 ~ 5,7%

In tal caso, per la categoria "operai" gli stipendi hanno una concentrazione più bassa (0,057) di quella rilevata per gli "impiegati" (0,12).

b)

Per costruire il diagramma di Lorenz basta porre sugli assi cartesiani le quantità p_i e q_i determinate precedentemente (e riportate nelle tabelle accanto alle distribuzioni di frequenze):



Come già dedotto dai valori del rapporto R calcolato precedentemente per i due gruppi, la concentrazione degli stipendi per gli impiegati è più alta che per gli operai. Infatti, la spezzata di concentrazione di questi ultimi (linea rossa) è più vicina alla retta di equidistribuzione rispetto a quella degli impiegati (linea blu).