

## Università di Cassino

### Esercitazione di Statistica 1 del 9 novembre 2007

Dott.ssa Paola Costantini

Considerando il DATASET DIPENDENTI:

ID	Stipendio percepito	Età	N. di anni di servizio	Qualifica funzionale	Regime di impiego	Genere	Stato Civile
1	2650	40	15	Operaio	Tempo pieno	M	Non coniugato
2	2600	43	5	Operaio	Part time	F	Vedovo
3	2050	35	6	Impiegato	Tempo pieno	F	Coniugato
4	3500	27	6	Dirigente	Part time	M	Non coniugato
5	1400	36	3	Dirigente	Collaboratori esterni	F	Vedovo
6	2400	30	12	Impiegato	Tempo pieno	M	Vedovo
7	1900	41	13	Operaio	Tempo pieno	F	Non coniugato
8	2100	35	4	Impiegato	Tempo pieno	M	Vedovo
9	2100	27	7	Operaio	Tempo pieno	F	Non coniugato
10	3050	38	18	Dirigente	Tempo pieno	F	Coniugato
11	2800	38	20	Operaio	Collaboratori esterni	M	Non coniugato
12	2950	41	11	Operaio	Collaboratori esterni	F	Non coniugato
13	1900	36	4	Dirigente	Collaboratori esterni	M	Vedovo
14	1650	29	11	Impiegato	Collaboratori esterni	F	Coniugato
15	2550	40	4	Impiegato	Collaboratori esterni	M	Non coniugato
16	2000	23	10	Impiegato	Tempo pieno	F	Coniugato
17	2150	26	8	Operaio	Collaboratori esterni	F	Coniugato
18	2900	41	9	Dirigente	Tempo pieno	M	Non coniugato
19	2450	35	12	Operaio	Collaboratori esterni	F	Coniugato
20	1950	31	8	Dirigente	Collaboratori esterni	M	Vedovo

1) Calcolare l'indice di eterogeneità di Gini per i caratteri "Qualifica Funzionale" e "Regime di Impiego".

#### **Condizione di massima omogeneità:**

le  $n$  unità statistiche presentano tutte la stessa modalità

#### **Condizione di massima eterogeneità:**

nella distribuzione di frequenza appaiono tutte le  $k$  modalità, e ad ognuna di esse è associata la medesima frequenza.

1.1) Calcoliamo l'indice del Gini per il carattere Qualifica Funzionale

Qualifica funzionale	ni	f <sub>i</sub>
operaio	8	0,4
impiegato	6	0,3
dirigente	6	0,3
<b>Totale</b>	<b>20</b>	<b>1,00</b>

$$G = 1 - \sum f_i^2 = 1 - (0,4^2 + 0,3^2 + 0,3^2) = 1 - (0,16 + 0,09 + 0,09) = 0,66$$

$$G_{max} = 1 - 1/k = 1 - 1/3 = 0,666$$

$$G^* = G / G_{max} = 0,66 / 0,67 = \mathbf{0,99}$$

### Conclusione

**G\* molto prossimo ad 1 → la distribuzione è molto eterogenea: tutte le modalità sono presenti e con frequenze molto simili tra loro.**

1.2) Calcoliamo l'indice del Gini per il carattere Regime di Impiego

Regime di impiego	ni	f <sub>i</sub>
Tempo pieno	9	0,45
Part time	2	0,10
Collaboratore esterno	9	0,45
<b>Totale</b>	<b>20</b>	<b>1,00</b>

$$G = 1 - \sum f_i^2 = 1 - (0,45^2 + 0,1^2 + 0,45^2) = 1 - (0,2025 + 0,01 + 0,2025) = 0,585$$

$$G_{max} = 1 - 1/k = 1 - 1/3 = 0,666$$

$$G^* = G / G_{max} = 0,585 / 0,67 = \mathbf{0,8775}$$

**G\* è abbastanza elevato → la distribuzione è abbastanza eterogenea: entrambe le modalità sono presenti, ma con frequenze non equilibrate tra loro.**

2) Suddividere in tre classi equiampie il carattere Stipendio percepito del dataset *Dipendenti*.

Anzitutto ordiamo le modalità del carattere Stipendio percepito

Stipendio Percepito
1400
1650
1900
1900
1950
2000
2050
2100
2100
2150
2400
2450
2550
2600
2650
2800
2900
2950
3050
3500

L'ampiezza costante delle tre classi si ottiene come:

$$\text{range (STIPENDIO)} = 3500 - 1400 = 2100$$

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

$$\text{Ampiezza **stipendio percepito**} = \text{range}/3 = 2100/3 = 700$$

Le 3 classi sono, dunque, delimitate dai seguenti estremi:

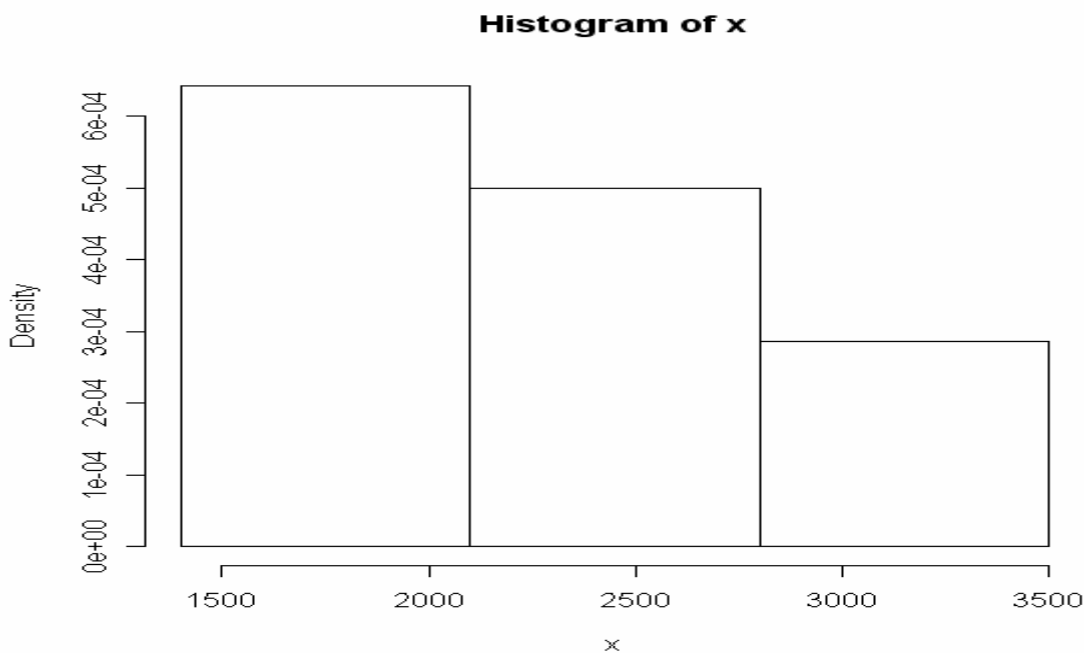
$$C_1 = [1400; 2100 ]$$

$$C_2 = ]2100; 2800 ]$$

$$C_3 = ] 2800; 3500 ]$$

$x_i$	$c_i$	$n_i$	$f_i$	$N_i$	$F_i$	$a_i$	$d_i$
[1400; 2100 ]	1750	9	0,45	9	0,45	700	0,00064
]2100; 2800 ]	2450	7	0,35	16	0,80	700	0,0005
] 2800; 3500 ]	3150	4	0,20	20	1	700	0,00028
<b>Totali</b>		<b>20</b>	<b>1,00</b>				

2.1) rappresentare graficamente tale distribuzione:



2.2) Calcolare, per la distribuzione in classi di cui sopra, gli **indici di posizione**: media, classe modale, mediana, quartili, gli **indici di variabilità**: varianza, scarto quadratico medio, coefficiente di variazione, gli **indici di posizione**: indice di Yule-Bowley e indice di Hotelling-Solomon.

### Indici di posizione

$$\text{Media} = \mu_x = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = 2275$$

**Classe modale** = [1440-2100]

Per il calcolo dei **Quartili** e della **Mediana**:

- 1) si individua la classe di riferimento;
- 2) si sostituisce ad  $F_{\text{desiderata}}$  il valore 0,25 per  $Q_1$ , 0,5 per  $Q_2$  e 0,75 per  $Q_3$ .

In base alla frequenza relativa cumulata, il **primo quartile** si trova nella classe **[1400 - 2100]**, quindi:

$$Q_1 = 1400 + (2100 - 1400) \frac{0,25 - 0}{0,45 - 0} = 1788,89$$

In base alla frequenza relativa cumulata, la **mediana** si trova nella classe **], quindi:**

$$\mathbf{Me} = 2100 + (2800 - 2100) \frac{0,5 - 0,45}{0,80 - 0,45} = \mathbf{2200}$$

In base alla frequenza relativa cumulata, il **terzo quartile** si trova nella classe **], quindi:**

$$\mathbf{Q_3} = 2100 + (2800 - 2100) \frac{0,75 - 0,45}{0,80 - 0,45} = \mathbf{2700}$$

### Indici di variabilità

per una distribuzione in classi, la **varianza** si calcola =  $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^k (c_i - \mu)^2 n_i$

dove  $c_i$  rappresenta il valore centrale della classe:

$$\sigma^2 = \frac{(1750 - 2275)^2 * 9 + (2450 - 2275)^2 * 7 + (3150 - 2275)^2 * 4}{20}$$

$$= \frac{2480625 + 214375 + 3062500}{20} = \mathbf{287.875}$$

### Scarto quadratico medio

$$\sigma = \sqrt{\sigma^2} = \sqrt{287,875} = \mathbf{536,5}$$

### Coeff. variazione =

$$CV = \frac{\sigma}{\mu} = \frac{536,5}{2275} = \mathbf{0,235}$$

### Indici di forma

<b>Yule-Bowley</b> =	$\frac{2Me - Q_3 - Q_1}{Q_3 - Q_1}$	$A_{YB} = 0$	Simmetria
		$A_{YB} < 0$	Asimmetria Positiva
		$A_{YB} > 0$	Asimmetria Negativa

$$\text{Yule-Bowley} = \frac{2*2200-2700-1788,89}{2700-1788,89} = -0,097 \quad \text{Asimmetria positiva}$$

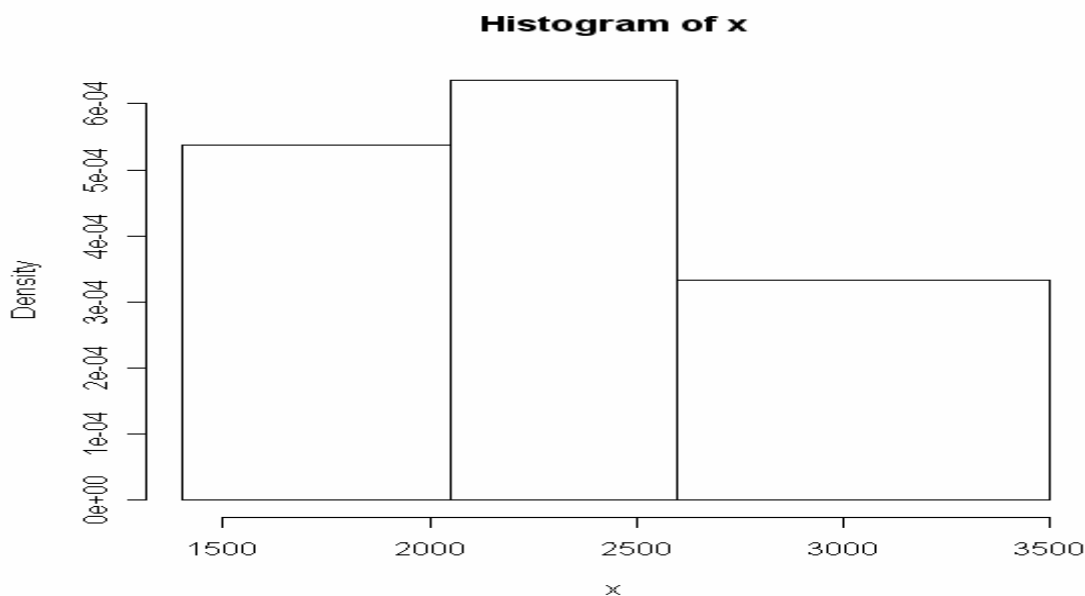
$$\text{Hotelling-Solomon} = \frac{\mu_x - Me}{\sigma} = \frac{2275 - 2200}{536,5} = 0,139 = \text{Asimmetria positiva}$$

3) Per lo stesso carattere Stipendio Percepito costruire una distribuzione con tre classi equifrequenti:

La numerosità del carattere Stipendio percepito è 20, quindi la frequenza costante nelle 3 classi dovrebbe essere pari a  $20/3 = 6,6667$ , che non ha senso ma suggerisce di scegliere una frequenza pari a 6 o a 7. Considerando anche l'ampiezza delle classi decidiamo:

Stipendio percepito	ci	n <sub>i</sub>	f <sub>i</sub>	N <sub>i</sub>	F <sub>i</sub>	a <sub>i</sub>	d <sub>i</sub>
[1400; 2050 ]	1725	7	0,35	7	0,35	650	0,00053
]2050; 2600 ]	2350	7	0,35	14	0,70	500	0,00070
]2600; 3500 ]	3075	6	0,30	20	1	900	0,00033
<b>Totali</b>		<b>20</b>	<b>1,00</b>				

3.1 Rappresentare graficamente tale distribuzione:



3.2) Calcolare, per la distribuzione in classi di cui sopra, gli **indici di posizione**: media, classe modale, mediana, quartili, gli **indici di variabilità**: varianza, scarto quadratico medio, coefficiente di variazione, gli **indici di posizione**: indice di Yule-Bowley e indice di Hotelling-Solomon.

### Indici di posizione

$$\text{Media} = \mu_x = \frac{\sum_{i=1}^k x_i \cdot n_i}{n} = 2348,75$$

**Classe modale** = [2050-2600] quella con la densità di frequenza più elevata

Per il calcolo dei **Quartili** e della **Mediana**:

3) si individua la classe di riferimento;

4) si sostituisce ad  $F_{\text{desiderata}}$  il valore 0,25 per  $Q_1$ , 0,5 per  $Q_2$  e 0,75 per  $Q_3$ .

In base alla frequenza relativa cumulata, il **primo quartile** si trova nella classe **[1400 - 2050]**, quindi:

$$Q_1 = 1400 + (2050 - 1400) \frac{0,25 - 0}{0,35 - 0} = 1864,1$$

In base alla frequenza relativa cumulata, la **mediana** si trova nella classe **]2050 - 2600]**, quindi:

$$Me = 2050 + (2600 - 2050) \frac{0,5 - 0,35}{0,70 - 0,35} = 2285,7$$

In base alla frequenza relativa cumulata, il **terzo quartile** si trova nella classe **]2600 - 3500]**, quindi:

$$Q_3 = 2600 + (3500 - 2600) \frac{0,75 - 0,70}{1 - 0,70} = 2750$$

## Indici di variabilità

per una distribuzione in classi, la **varianza** si calcola =  $\sigma_x^2 = \frac{1}{n} \sum_{i=1}^k (c_i - \mu)^2 n_i$

dove  $c_i$  rappresenta il valore centrale della classe:

$$\sigma^2 = \frac{(1725-2348,75)^2 * 7 + (2350-2348,75)^2 * 7 + (3075-2348,75)^2 * 6}{20} =$$

$$= \frac{2723448,4 + 10,93 + 3164634,3}{20} = \mathbf{294404,7}$$

## Scarto quadratico medio

$$\sigma = \sqrt{\sigma^2} = \sqrt{294404,7} = \mathbf{542,6}$$

## Coeff. variazione =

$$CV = \frac{\sigma}{\mu} = \frac{542,6}{2348,75} = \mathbf{0,231}$$

## Indici di forma

$$\text{Yule-Bowley} = \frac{2Me - Q_3 - Q_1}{Q_3 - Q_1}$$

$A_{yB} = 0$	Simmetria
$A_{yB} < 0$	Asimmetria Positiva
$A_{yB} > 0$	Asimmetria Negativa

$$\text{Yule-Bowley} = \frac{2 * 2285,7 - 2750 - 1864,1}{2750 - 1864,1} = \mathbf{-0,0482} \quad \underline{\underline{\text{Asimmetria positiva}}}$$

$$\text{Hotelling-Solomon} = \frac{\mu_x - Me}{\sigma} = \frac{2348,75 - 2285,7}{542,6} = 0,116 = \mathbf{\text{Asimmetria positiva}}$$