

# CORSO DI STATISTICA 1

## Esercitazione N°2

(28 gennaio 2011)

Dr. Mirko Bevilacqua

### Dataset "test d'ingresso"

	SEDE DISTACCATA	GENERE	VOTO DIPLOMA	TIPO SCUOLA	PROVINCIA SEDE SCUOLA	CITTA SCUOLA	PUNTEGGIO TEST
1	SEDE DI CASSINO	F	60	TC	FR	CASSINO	5
2	SEDE DI TERRACINA	M	60	LC	LT	FONDI	14,75
3	SEDE DI CASSINO	F	69	LS	FR	CECCANO	15
4	SEDE DI TERRACINA	M	100	GE	LT	FORMIA	15
5	SEDE DI CASSINO	M	73	TC	FR	CASSINO	17,25
6	SEDE DI CASSINO	M	70	LS	FR	VEROLI	17,5
7	SEDE DI CASSINO	F	72	MP	FR	CASSINO	8,75
8	SEDE DI CASSINO	M	64	LS	FR	PONTECORVO	17,25
9	SEDE DI TERRACINA	M	72	LS	FR	CECCANO	18,5
10	SEDE DI TERRACINA	M	70	TC	FR	CECCANO	9,25
11	SEDE DI CASSINO	M	63	LS	FR	SORA	15,75
12	SEDE DI CASSINO	F	76	MP	FR	CASSINO	7,25
13	SEDE DI CASSINO	F	80	TC	FR	CASSINO	8,5
14	SEDE DI TERRACINA	M	100	TC	LT	FORMIA	26,15
15	SEDE DI CASSINO	M	73	TC	LT	FORMIA	22,3
16	SEDE DI CASSINO	F	73	LS	LT	FORMIA	28,8

### ESERCIZIO N1

Considerando il dataset "test d'ingresso" costruire le distribuzioni doppie di frequenza per le seguenti coppie di caratteri: "sesso"- "Tipo scuola" e "Sesso Voto diploma".

A)

Sesso \ Tipo Sc.	Sesso		Totale
	M	F	
GE	1	0	1
LC	1	0	1
LS	4	2	6
MP	0	2	2
TC	4	2	6
Totale	10	6	16

B)

Voto D.	Sesso		Totale
	M	F	
60	1	1	2
63	1	0	1
64	1	0	1
69	0	1	1
70	2	0	2
72	1	1	2
73	2	1	3
76	0	1	1
80	0	1	1
100	2	0	2
<b>Totale</b>	<b>10</b>	<b>6</b>	<b>16</b>

## ESERCIZIO N 2

Calcolare media aritmetica, moda e mediana e quartili per il carattere VOTO DIPLOMA a partire:

**2.1)** dalla distribuzione semplice del carattere;

**2.2)** dalla distribuzione nelle classi: [60–70]; ]70–80]; ]80–90]; ]90–100].

### 2.1

#### Media aritmetica

Voto Dip.	60	63	64	69	70	72	73	76	80	100
$n_i$	2	1	1	1	2	2	3	1	1	2

$$\mu_{VD} = \frac{\sum_{i=1}^n x_i \cdot n_i}{n} = \frac{\sum_{i=1}^{16} x_i \cdot n_i}{16} = \frac{60 \cdot 2 + 63 + \dots + 80 + 100 \cdot 2}{16} = 73,4$$

#### Moda

La **Moda** è la modalità della distribuzione che si presenta con la massima frequenza

Osservando la distribuzione di frequenze:

Voto Dip.	60	63	64	69	70	72	73	76	80	100
$n_i$	2	1	1	1	2	2	3	1	1	2

$$\text{Moda}_{(\text{voto diploma})} = 73$$

## Mediana

La Mediana (Me) di un insieme di unità ordinate (secondo un carattere ordinabile) è la modalità presentata dall'unità centrale dove per unità centrale si intende quell'unità che divide il collettivo in due parti di uguale numerosità.

60	60	63	64	69	70	70	72	72	73	73	73	76	80	100	100
----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

$$Q_2 = Me = \frac{x_{\frac{N}{2}} + x_{\frac{N}{2}+1}}{2} = \frac{x_8 + x_9}{2} = \frac{72 + 72}{2} = 72$$

## Quartili

I quartili della distribuzione semplice vanno individuati determinando prima la mediana (secondo quartile) e poi le mediane delle due metà della distribuzione che si trovano a sinistra (primo quartile) e a destra (terzo quartile) della mediana.

### Primo quartile

$$Q_1 = \frac{x_4 + x_5}{2} = \frac{64 + 69}{2} = 66,5$$

60	60	63	64	69	70	70	72	72	73	73	73	76	80	100	100
----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

### Terzo quartile

$$Q_3 = \frac{x_{12} + x_{13}}{2} = \frac{73 + 76}{2} = 74,5$$

60	60	63	64	69	70	70	72	72	73	73	73	76	80	100	100
----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----

## 2.2

$x_{i-1} - x_i$	$c_i$	$n_i$	$f_i$	$F_i$	$a_i$	$d_i$
[60 - 70]	65	7	0,44	0,44	11	0,04
]70 - 80]	75	7	0,44	0,88	10	0,04
]80 - 90]	85	0		0,88	10	0,04
]90 - 100]	95	2	0,12	1	10	0,01
<b>Totali</b>		<b>16</b>	<b>1</b>			

## Media

Data la distribuzione di frequenza del carattere altezza suddivisa in 4 classi equiampie si può approssimare la Media aritmetica del carattere con l'espressione riportata sotto, dove  $k$  è il numero di classi della distribuzione di frequenza,  $c_i$  è il valore centrale della classe  $i$ -esima (il valore che si ottiene come semisomma degli estremi della classe) e  $n_i$  è la corrispondente frequenza assoluta:

$$\mu_{VD} = \frac{\sum_{i=1}^k C_i \cdot n_i}{n} = \frac{\sum_{i=1}^4 C_i \cdot n_i}{16} =$$

$$= \frac{(65 \times 7) + (75 \times 7) + (95 \times 2)}{16} = 73,2$$

### Classe Modale

La Classe modale è definita come la classe alla quale corrisponde la frequenza più alta. Se la distribuzione possiede classi di diversa ampiezza occorre dividere le frequenze delle classi per la loro ampiezza e confrontare tali quozienti: quello più grande individuerà la classe modale.

classe modale = **[60 – 70]**

### Mediana

La classe mediana è ]70-80]: prima classe avente frequenza relativa cumulata maggiore o uguale a 0,5. Individuata la classe mediana, si procede con la stima della mediana all'interno della classe, utilizzando la seguente formula:

$$Me = x_{Me-1} + \frac{(0,5 - F_{Me-1})}{F_{Me} - F_{Me-1}} \cdot (x_{Me} - x_{Me-1}) \quad \text{dove}$$

**$x_{Me-1}$**  e  **$x_{Me}$**  individuano l'estremo inferiore e l'estremo superiore della classe dove cade la mediana

**$F_{Me-1}$**  e  **$F_{Me}$**  indicano la frequenza relativa cumulata fino alla classe precedente a quella in cui cade la mediana e la frequenza relativa cumulata della classe mediana.

$$Q_2 = Me = 70 + \frac{(0,5 - 0,44)}{(0,88 - 0,44)} \cdot (80 - 70) = 170$$

### Primo quartile

Il procedimento è analogo al precedente. Naturalmente la classe di riferimento è la prima che presenta frequenza relativa cumulata maggiore o uguale a 0,25.

- Classe di riferimento: [60-70].

$$Q_1 = x_{Q_1-1} + \frac{(0,25 - F_{Q_1-1})}{(F_{Q_1} - F_{Q_1-1})} \cdot (x_{Q_1} - x_{Q_1-1}) = 60 + \frac{0,25 - 0,44}{0,88 - 0,44} \cdot (70 - 60) = 65,7$$

### Terzo quartile

Stesso metodo, ma con frequenza relativa cumulata maggiore o uguale a 0,75.

Classe di riferimento: ]70-80].

$$Q_3 = x_{Q_3-1} + \frac{(0,75 - F_{Q_3-1})}{(F_{Q_3} - F_{Q_3-1})} \cdot (x_{Q_3} - x_{Q_3-1}) = 70 + \frac{(0,75 - 0,44)}{(0,88 - 0,44)} \cdot (80 - 70) = 77,0$$

## REGOLA GENERALE

I quantili di una distribuzione in classi possono essere determinati in base alla formula generica:

$$x_{P_X} = x_{P_X-1} + \frac{(F_{\text{desiderata}} - F_{P_X-1})}{F_{P_X} - F_{P_X-1}} \cdot (x_{P_X} - x_{P_X-1})$$

## ESERCIZIO N3

**3.1)** Verificare la proprietà di linearità della media aritmetica del carattere "punteggio test", ipotizzando un aumento di 1 punto del punteggio finale (punteggio test) per tutti i partecipanti al test d'ingresso.

### 3.1

$$\mu_{PT} = \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^{16} x_i}{16} = \frac{5 + 14,75 + \dots + 22,3 + 28,8}{16} = 15,4$$

La media aritmetica di un carattere Y ottenuto attraverso una trasformazione lineare  $Y=a+bX$  di un carattere X di media  $\mu_x$  è uguale a  $\mu_y = a+b\mu_x$ .

$$a=1; \quad b=1;$$

$$\mu_{PT+1} = 1 + \mu_{PT} = 15,4 + 1 = 16,4$$

Ricalcolando la media aritmetica aumentando di una unità il punteggio di tutti gli studenti si ha:

$$\frac{\sum_{i=1}^n (x_i + 1)}{n} = \frac{\sum_{i=1}^{16} (x_i + 1)}{16} = \frac{6 + 15,75 + \dots + 23,3 + 29,8}{16} = 16,4$$

**3.2)** Verificare la proprietà associativa della media aritmetica del carattere "Punteggio test" rispetto al carattere "Tipo scuola".

### 3.2

Se un collettivo di n unità statistiche viene suddiviso in L sottoinsiemi disgiunti di numerosità  $n_1, n_2, n_3, \dots, n_L$ , allora la media aritmetica generale si può ottenere come media ponderata delle medie dei sottoinsiemi con pesi uguali alla loro numerosità.

Di seguito le medie rispetto del carattere punteggio test rispetto alle singole modalità della variabile tipo di scuola:

$$\mu_{PT/TC} = 14,7 \quad - \quad \mu_{PT/GE} = 15 \quad - \quad \mu_{PT/LC} = 14,75 \quad - \quad \mu_{PT/LS} = 18,8 \quad - \quad \mu_{PT/MP} = 8$$

$$\begin{aligned} \mu_{PT} &= \frac{\mu_{PT/TC} \cdot n_{TC} + \mu_{PT/GE} \cdot n_{GE} + \mu_{PT/LC} \cdot n_{LC} + \mu_{PT/LS} \cdot n_{LS} + \mu_{PT/MP} \cdot n_{MP}}{n} \\ &= \frac{14,7 \cdot 6 + 15 + 14,75 + 18,8 \cdot 6 + 8 \cdot 2}{16} = 15,4 \end{aligned}$$

## ESERCIZIO N 4

Rappresentare con un istogramma la distribuzione in classi del carattere VOTO DIPLOMA dopo averla ripartita in 5 classi equiampie. Calcolare la media aritmetica e gli altri indici di posizione.

4.

*Campo di Variazione della variabile = 100-60=40*  
*ampiezza delle classi = 40:5=8*

$x_{i-1} - x_i$	$c_i$	$n_i$	$f_i$	$F_i$	$a_i$	$d_i$
[60 - 68]	64	4	0,25	0,25	9	0,03
]68 - 70]	72	9	0,56	0,81	8	0,07
]76 - 84]	80	1	0,06	0,87	8	0,01
]84 - 92]	88	0	0	0,87	8	0,00
]92 -100]	96	2	0,13	1	8	0,02
<b>Totali</b>		<b>16</b>	<b>1</b>			

### Media aritmetica

$$\mu_{VD} = \frac{\sum_{i=1}^k C_i \cdot n_i}{n} = \frac{\sum_{i=1}^4 C_i \cdot n_i}{16} =$$
$$= \frac{(65 \times 7) + (75 \times 7) + (95 \times 2)}{16} = 73,2$$

### Classe Modale

classe modale = **[68 - 76]**

### Mediana

$$Me = x_{Me-1} + \frac{(0,5 - F_{Me-1})}{F_{Me} - F_{Me-1}} \cdot (x_{Me} - x_{Me-1}) = 68 + \frac{(0,5 - 0,25)}{0,81 - 0,25} \cdot (76 - 68) = 71,57$$

### Primo quartile

$$Q_1 = x_{Q_1-1} + \frac{(0,25 - F_{Q_1-1})}{(F_{Q_1} - F_{Q_1-1})} \cdot (x_{Q_1} - x_{Q_1-1}) = 60 + \frac{0,25 - 0}{0,25 - 0} \cdot (68 - 60) = 68$$

### Terzo quartile

$$Q_3 = x_{Q_3-1} + \frac{(0,75 - F_{Q_3-1})}{(F_{Q_3} - F_{Q_3-1})} \cdot (x_{Q_3} - x_{Q_3-1}) = 68 + \frac{(0,75 - 0,25)}{(0,81 - 0,25)} \cdot (76 - 68) = 76,96$$

Histogram of VotoDiploma

