

# Corso di Statistica

## Esercizio 1

Nella seguente tabella sono riportate le distribuzioni unitarie riguardanti la provenienza (Italiani, Stranieri) e la durata del soggiorno (in giorni) di un collettivo di 30 turisti:

<b>Provenienza (Y)</b>	S	S	I	S	S	S	S	I	I	S	I	I	I	I	S	I	S	S	S	S	I	I	I	I	I	I	S	I	I	I	
<b>Durata (X)</b>	3	6	3	5	5	7	5	2	7	4	1	5	2	3	3	2	6	6	4	5	1	4	2	2	2	2	1	7	3	2	4

Descrivere statisticamente attraverso gli indici di posizione e variabilità più opportuni, entrambe le distribuzioni e rappresentarle graficamente.

\*\*\*\*\*

Distribuzione **durata** del soggiorno (x).

DURATA $x_i$	$n_i$	$N_i$	$f_i$	$F_i$	$x_i * n_i$	$x_i^2$	$x_i^2 * n_i$
1	3	3	0,10	0,10	3	1	3
2	7	10	0,23	0,33	14	4	28
3	5	15	0,17	0,50	15	9	45
4	4	19	0,13	0,63	16	16	64
5	5	24	0,17	0,80	25	25	125
6	3	27	0,10	0,90	18	36	108
7	3	30	0,10	1,00	21	49	147
<b>totale</b>	<b>30</b>		<b>1,00</b>		<b>112</b>		<b>520</b>

Calcolo della **media**:

per una distribuzione di frequenza semplice è:  $\mu = \frac{1}{N} \sum_{i=1}^k x_i n_i$  oppure  $\mu = \sum_{i=1}^k x_i f_i$

$$\text{quindi: } \mu = \frac{112}{30} = 3,73$$

Calcolo della **varianza** con la formula abbreviata:  $\sigma^2 = \frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \mu^2 = \frac{520}{30} - 3,73^2 = 3,40$

Calcolo della **deviazione standard**:  $\sigma = \sqrt{\frac{\sum_{i=1}^k x_i^2 * n_i}{N} - \mu^2} = \sqrt{3,40} = 1,84$ .

Calcolo del **coefficiente di variazione**:  $CV = \frac{\sigma}{\mu} = \frac{1,84}{3,73}$

Calcolo della **mediana** (N pari):  $Me = \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2} = \frac{x_{(15)} + x_{(16)}}{2} = \frac{3+4}{2} = 3,5$

La **moda** è 2, poiché risulta essere l'intensità più frequente (7).

Rappresentiamo graficamente la nostra distribuzione attraverso il **BOX PLOT**

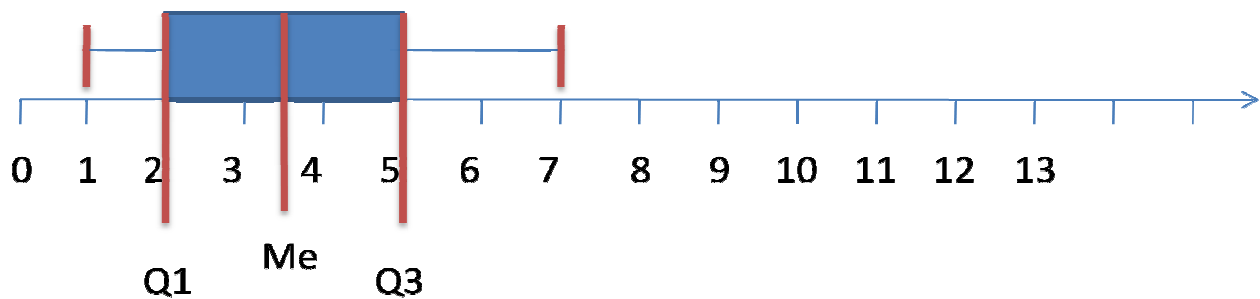
Calcolo dei **quartili**:  $Q_1 = x_{(\frac{N}{4})} = x_{(8)} = 2$        $Q_3 = x_{(\frac{N}{4} * 3)} = x_{(23)} = 5$

Differenza interquartile =  $(Q_3 - Q_1) = 5 - 2 = 3$

**Baffo inferiore** =  $Q_1 - 1,5(Q_3 - Q_1) = 2 - 1,5(3) = -2,5 = 0$ . Poiché non abbiamo tale valore prendiamo quello più vicino al limite inferiore che coincide proprio col minimo della distribuzione (1), inoltre, non ha senso una durata del soggiorno negativa.

**Baffo superiore** =  $Q_3 + 1,5(Q_3 - Q_1) = 5 + 1,5(3) = 9,5$ . Poiché non abbiamo tale valore prendiamo quello più vicino al limite superiore che coincide proprio col massimo della distribuzione (7).

**BOX-PLOT**



Distribuzione **provenienza** (y).

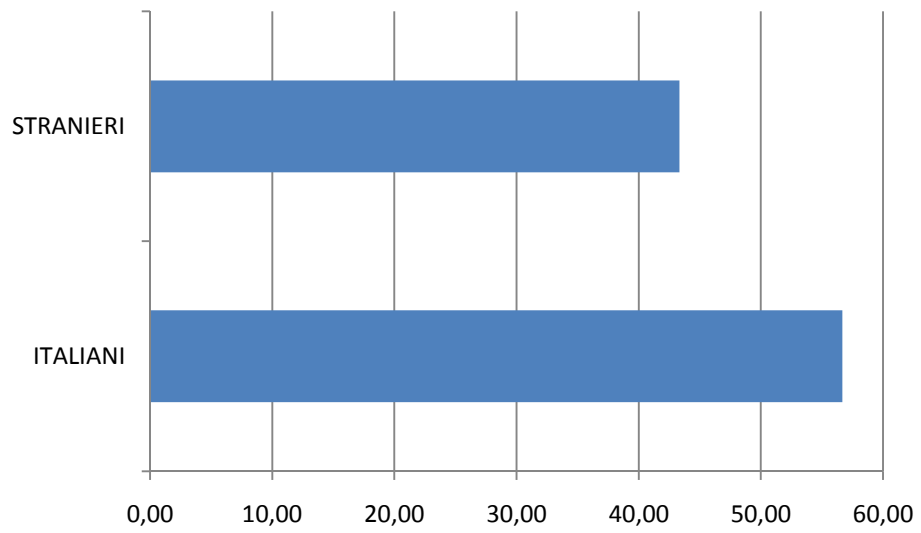
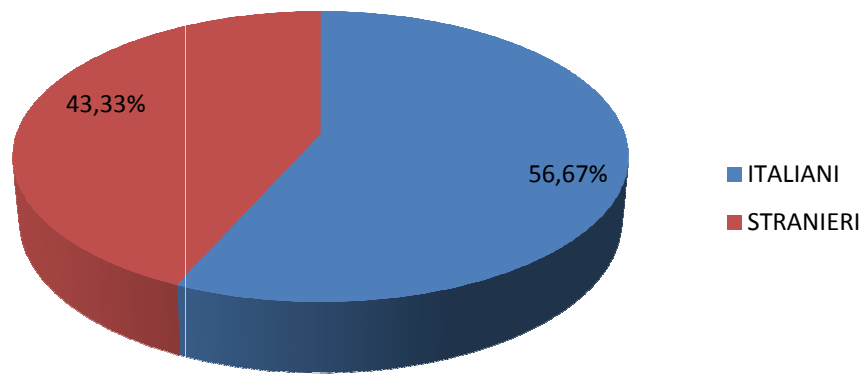
Essendo la provenienza un carattere qualitativo sconnesso possiamo calcolare la **Moda** che è pari alla modalità Italiani, poiché è la più frequente.

PROVENIENZA $y_i$	$n_i$	$f_i$	$f_i * 100$
ITALIANI	17	0,567	56,67
STRANIERI	13	0,433	43,33
TOTALE	30	1,00	100

Si può rappresentare graficamente la distribuzione unitaria del carattere qualitativo sconnesso “Provenienza”.

Esempio tipico di rappresentazione grafica per tale tipo di variabile è il **grafico a torta**. Si tratta di un grafico di composizione, il cui scopo è mettere in evidenza la ripartizione di un dato globale nei dati parziali che lo compongono. In altre parole, questo tipo di grafico pone in risalto la “struttura” del fenomeno considerato.

Altra possibile rappresentazione grafica utilizzabile nel caso di carattere qualitativo sconnesso è il **grafico a barre orizzontali**. In tale tipo di grafico la frequenza di ogni modalità è rappresentata da un segmento di larghezza fissa e di lunghezza proporzionale alla frequenza della modalità stessa. Questo tipo di grafico viene utilizzato, inoltre, nel caso sia necessario confrontare due o più distribuzioni riferite a collettivi distinti, ad esempio la distribuzione delle spese per tipo di spettacoli in città diverse.



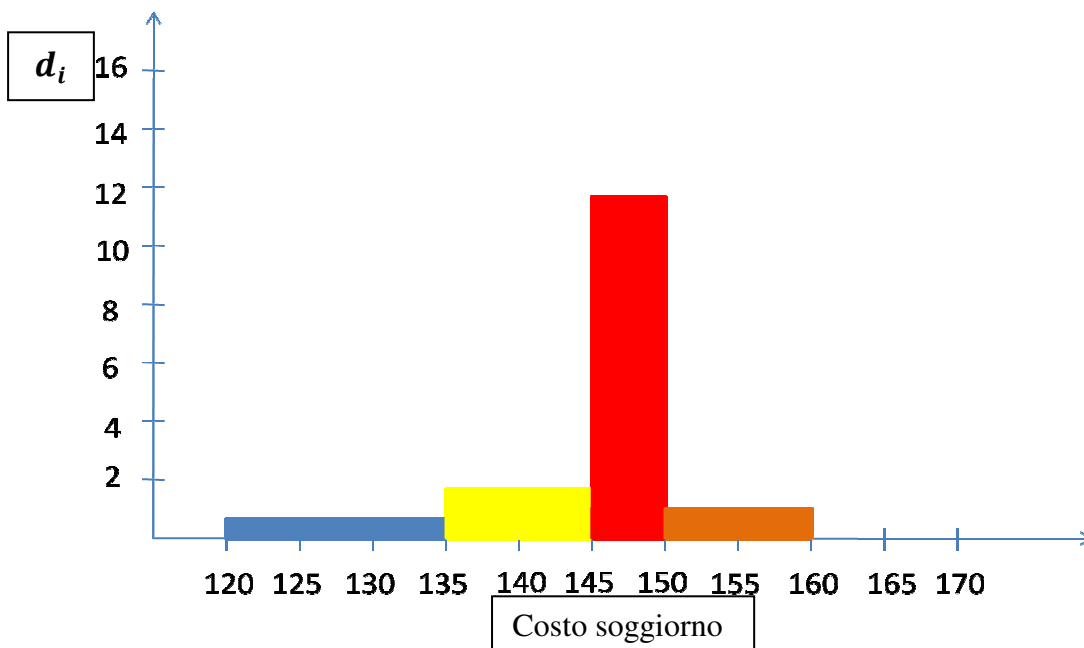
## Esercizio 2

Di seguito si riporta la distribuzione del costo del soggiorno sulla base dei prezzi praticati da 100 albergatori:

<b>Costo soggiorno</b>	$n_i$	$a_i$	$d_i$	$c_i$	$c_i * n_i$	$c_i^2$	$c_i^2 * n_i$
Da 120 a 135	10	15	0,67	127,5	1275	16256,25	162562,5
Da 135 a 145	20	10	2	140	2800	19600,00	392000
Da 145 a 150	60	5	12	147,5	8850	21756,25	1305375
Da 150 a 160	10	10	1	155	1550	24025,00	240250
<b>totale</b>	<b>100</b>				<b>14475</b>	<b>81637,5</b>	<b>2100188</b>

E' noto che nelle festività natalizie gli albergatori applicano un incremento del 15%.

- a) Si rappresentino le osservazioni tramite un istogramma e si calcoli media e varianza;
- b) Si calcolino la media e la varianza nel periodo natalizio.



**Calcoliamo la media:** 
$$\mu = \frac{1}{N} \sum_{i=1}^k c_i n_i = \frac{14475}{100} = 144,75$$

Il quadrato della media:  $\mu^2 = 144,75^2 = 20952,56$

**Calcoliamo la varianza:**

$$\sigma^2 = \frac{\sum_{i=1}^k c_i^2 * n_i}{N} - \mu^2 = \frac{2100188}{100} - 20952,56 = 21001,88 - 20952,56 = 49,32$$

**Calcoliamo la media e la varianza nel periodo natalizio:**

Date le proprietà della media e della varianza basta moltiplicare i valori ottenuti in precedenza per la costante 1,15:

**media:**  $\mu = 144,75 * 1,15 = 166,46$

**varianza:**  $\sigma^2 = (49,32 * 1,15^2) = 65,23$