RAPPORTI STATISTICI

La seguente tabella riporta il numero di studenti immatricolati all'Università di Bari nell'a.a. 2013/14 per provincia di provenienza e per Dipartimento.

PROVENIENZA	Scienze Politiche	Giurisprudenza	Scienze Economiche
Provincia di Bari	643	699	465
Altre province della regione	308	362	137
Altre regioni e estero	107	92	68
Totale	**1058**	**1153**	**670**

[0.1] Calcolare il Rapporto di composizione per gli immatricolati provenienti dalla provincia di Bari secondo il Dipartimento di afferenza e commentare i risultati.

[0.2] Determinare l'incidenza percentuale degli immatricolati provenienti da altre regioni e dall'estero nel Dipartimento di Scienze Politiche.

[0.2] SOLUZIONE

Il *Rapporto di composizione* si ottiene riportando la frequenza di una parte del fenomeno alla frequenza complessiva del medesimo ("al tutto" del carattere).

PROVENIENZA	Scienze Politiche	Giurisprudenza	Scienze economiche
Provincia di Bari	(643/1058)*100= 60,8%	(699/1153)*100= 60,6%	(465/670)*100= 69,4%
Altre province della regione			
Altre regioni e estero			
Totale			

I Rapporti calcolati per Dipartimento mostrano che la quota di immatricolati provenienti dalla provincia di Bari è più elevata nel Dipartimento di Scienze economiche.

[0.2] SOLUZIONE

PROVENIENZA	Scienze Politiche	Giurisprudenza	Scienze economiche
Provincia di Bari	60,8%		
Altre province della regione	29,1%		
Altre regioni e estero	10,1%		
Totale	100%		

MEDIE DI CALCOLO E DI POSIZIONE

A 20 studenti dei corsi di Statistica del Dipartimento di Scienze Politiche dell'Università di Bari viene somministrato un questionario contenente quattro quesiti su età, voto di diploma, titolo di studio del padre e provincia di provenienza.
Si riportano le informazioni ottenute nella seguente tabella:

STUDENTE	ETÀ	VOTO DIPLOMA	TITOLO STUDIO DEL PADRE	PROVINCIA DI PROVENIENZA
1	19	90	Diploma superiore	Bari
2	19	88	Laurea	Brindisi
3	19	85	Licenza media	BAT
4	18	95	Dottorato	Bari
5	20	62	Diploma superiore	Bari
6	21	65	Diploma superiore	BAT
7	18	70	Laurea	BAT
8	19	75	Laurea	Bari
9	19	75	Laurea	Bari
10	20	72	Diploma superiore	Lecce
11	24	65	Dottorato	Bari
12	24	63	Diploma superiore	BAT
13	21	67	Laurea	Brindisi
14	19	75	Laurea	Lecce
15	19	74	Dottorato	Taranto
16	19	71	Diploma superiore	Bari
17	18	70	Laurea	Taranto
18	19	70	Diploma superiore	Bari
19	20	68	Licenza media	Brindisi
20	19	80	Diploma superiore	Bari

[0.3] A partire dalla matrice dei dati costruire le distribuzioni delle frequenze assolute e relative per i caratteri, *età* e *voto di diploma*, utilizzando per il *voto di diploma* le seguenti classi: 60┤ 70, 70┤ 80, 80┤ 90, 90┤ 100.

[0.4] Calcolare le medie di calcolo (media aritmetica, armonica e quadratica) per le variabili statistiche *età* e *voto di diploma*.

[0.5] Calcolare le medie di posizione (moda, mediana e quartili) per le variabili statistiche *età* e *voto di diploma*.

[0.6] Per quale delle due variabili statistiche, *titolo di studio del padre* e *provincia di provenienza,* si possono calcolare la moda e la media?

[0.7] Calcolare la mediana per la variabile statistica: *titolo di studio del padre.*

[0.3] SOLUZIONE

$y_i = \frac{n_i}{N}$ *(frequenze relative);* $p_i = y_i * 100$ *(frequenze percentuali);*

N_i *(frequenze cumulate)*

x_i (età)	n_i	p_i	$p_i \cdot 100$	N_i
18	3	0,15	15	3
19	10	0,5	50	13
20	3	0,15	15	16
21	2	0,1	10	18
24	2	0,1	10	20
	N=20			

$x_i - x_{i+1}$(voto)	n_i	p_i	$p_i \cdot 100$	N_i
60⊢ 70	9	0,45	45	9
70⊢ 80	7	0,35	35	16
80⊢ 90	3	0,15	15	19
90⊢ 100	1	0,05	5	20
	N=20			

[0.4] SOLUZIONE

VARIABILE STATISTICA "ETÀ"

x_i (età)	n_i	$x_i \cdot n_i$	x_i^2	$x_i^2 \cdot n_i$	n_i / x_i
18	3	54	324	972	0,17
19	10	190	361	3610	0,53
20	3	60	400	1200	0,15
21	2	42	441	882	0,10
24	2	48	576	1152	0,08
	N=20	Σ=394	2102	Σ=7816	Σ=1,02

$$\overline{\chi} = \mu = \frac{\sum_{i=1}^{S} x_i \cdot n_i}{N} = \frac{394}{20} = 19{,}7; \qquad M_Q = \sqrt{\frac{\sum_{i=1}^{s} x_i^2 \cdot n_i}{N}} = \sqrt{\frac{7816}{20}} = 19{,}8;$$

$$Mar = \frac{N}{\sum_{i=1}^{s} \frac{n_i}{x_i}} = \frac{20}{1{,}02} = 19{,}60$$

VARIABILE STATISTICA "VOTO DI DIPLOMA"

Trattandosi di una variabile statistica divisa in classi, calcoliamo prima di tutto i valori centrali delle classi

$x_i - x_{i+1}$(voto)	n_i	v.c. (x_i)	$x_i \cdot n_i$	x_i^2	$x_i^2 \cdot n_i$	n_i / x_i	N_i
60-70	9	65	585	4225	38025	0,14	9
70-80	7	75	525	5625	39375	0,09	16
80-90	3	85	255	7225	21675	0,04	19
90-100	1	95	95	9025	9025	0,01	20
	N=20	320	Σ=1460	26100	Σ=108100	Σ=0,28	

$$\mu = \frac{\sum_{i=1}^{N} v.c._{x_i} \cdot n_i}{N} = \frac{1460}{20} = 73\ ; \quad M_Q = \sqrt{\frac{\sum_{i=1}^{s} x_i^2 \cdot n_i}{N}} = \sqrt{\frac{108100}{20}} = \sqrt{5405} = 73{,}51$$

$$Mar = \frac{N}{\sum_{i=1}^{s} \frac{n_i}{x_i}} = \frac{20}{0{,}28} = 72{,}04$$

[0.5] SOLUZIONE

VARIABILE STATISTICA "ETÀ":

1) MODA=19

2)

$$M_e = \chi_{\left(\frac{N}{2}\right)} + \chi_{\left(\frac{N+1}{2}\right)} = \frac{\chi_{(10)} + \chi_{(11)}}{2} = \frac{19+19}{2} = 19$$

$\chi_{(10)}$ cade in posizione $x = 19$; $\chi_{(11)}$ cade in posizione $x = 19$

3)

$$Q_1 = \frac{\chi_{\left(\frac{N}{4}\right)} + \chi_{\left(\frac{N}{4}+1\right)}}{2} = \frac{\chi_{(5)} + \chi_{(6)}}{2} = \frac{19+19}{2} = 19$$

$$Q_3 = \frac{\chi_{\left(\frac{3N}{4}\right)} + \chi_{\left(\frac{3N}{4}+1\right)}}{2} = \frac{\chi_{15} + \chi_{(16)}}{2} = \frac{20+20}{2} = 20$$

DIFFERENZA INTERQUARTILICA: $Q_3 - Q_1 = 20 - 19 = 1$

VARIABILE STATISTICA "VOTO DI DIPLOMA"

1) CLASSE MODALE= [60-70]

2)

$$M_e = \chi_{\left(\frac{N}{2}\right)} + \chi_{\left(\frac{N+1}{2}\right)} = \frac{\chi_{(10)} + \chi_{(11)}}{2} = \frac{19+19}{2} = [70-80]$$

Per le distribuzioni continue: $M_e = \chi_1 + \frac{\chi_{i+1} - \chi_i}{n_i}\left(\frac{N}{2} - N_{i-1}\right)$

Dove:
χ_1 è il primo estremo della classe mediana;
$\chi_{i+1} - \chi_i$ è l'ampiezza della classe;
n_i è la frequenza assoluta relativa alla classe mediana;
N_{i+1} è la frequenza cumulata precedente alla classe mediana.

$$M_e = 70 + \frac{80-70}{7}\left(\frac{20}{2} - 9\right) = 70 + \frac{10}{7}(10-9) = 70 + 1{,}42 = 71{,}42$$

3)

$$Q_1 = \frac{x_{\left(\frac{N}{4}\right)} + x_{\left(\frac{N}{4}+1\right)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = [60-70]$$

$$Q_i = \chi_1 + \frac{\chi_{i+1} - \chi_i}{n_i}\left(\frac{N}{4} - N_{i-1}\right) = 60 + \frac{10}{9}(5-0) = 60 + 5{,}55 = 65{,}55$$

$$Q_3 = \frac{x_{\left(\frac{3N}{4}\right)} + x_{\left(\frac{3N}{4}+1\right)}}{2} = \frac{x_{15} + x_{(16)}}{2} = [70 - 80]$$

$$Q3 = x_1 + \frac{x_{i+1} - x_i}{n_i}\left(\frac{3N}{4} - N_{i-1}\right) = 70 + \frac{60}{9}(15-9) = 70 + 8{,}57 = 78{,}57$$

[0.6] SOLUZIONE

Possiamo calcolare *la moda* per entrambe le due variabili statistiche:

- ✓ la moda della v.s. *titolo di studio del padre* corrisponde alla modalità "Diploma superiore" alla quale è associata la frequenza maggiore;
- ✓ -la moda della v.s *provincia di provenienza* corrisponde alla modalità "Bari" alla quale è associata la frequenza maggiore

Non possiamo calcolare *la media* perché le due variabili statistiche sono qualitative.

[0.7] SOLUZIONE

x_i *(titolo studio padre)*	n_i	y_i	N_i	Y_i
Licenza media	2	0,1	2	0,1
Diploma superiore	8	0,4	10	0,5
Laurea	7	0,35	17	0,85
Dottorato	3	0,15	20	1
	20	1		

La mediana corrisponde alla modalità la cui cumulata relativa è la prima a raggiungere e superare 0,5 (oppure il 50% nel caso di frequenza cumulata percentuale)

VARIABILITÀ

Nella seguente tabella è riportata la distribuzione della variabile statistica X= voto rilevata su 200 studenti che hanno sostenuto l'esame di statistica.

Voto x_i- x_{i+1}	*ni*
18-20	29
21-23	31
24-26	62
27-28	53
29-30	25
Totale	**200**

[0.1] Calcolare gli indici di variabilità secondo le classi di voto (scarto semplice medio, scarto quadratico medio, varianza, devianza).

[0.1] SOLUZIONE

.

x_i *(voto)*	n_i	*v.c.* (x_i)	$x_i \cdot n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x}) \cdot n_i$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$
18-20	29	19	551	-5,89	-170,81	34,69	1.006,07
21-23	31	22	682	-2,89	-89,59	8,35	258,92
24-26	62	25	1.550	0,11	6,82	0,01	0,75
27-28	53	27,5	1.457,5	2,61	138,33	6,81	361,04
29-30	25	29,5	737,5	4,61	115,25	21,25	531,30
	200		4.978				2.158,08

SCARTO SEMPLICE MEDIO:
Si procede al calcolo della *media aritmetica* applicando la formula:

$$\bar{x} = \mu = \frac{\sum_{i=1}^{S} x_i \cdot n_i}{N} = \frac{4.978}{200} = 24{,}89;$$

si esegue il rapporto tra la somma degli scarti in valore assoluto (ponderati per le rispettive frequenze) ed il collettivo N:

$$\delta = \frac{\sum_{i=1}^{S} |x_i - \mu| \cdot n_i}{N} = \frac{520{,}8}{200} = 2{,}6$$

SCARTO QUADRATICO MEDIO
Per calcolare lo *scarto quadratico medio* occorre eseguire la radice quadrata del rapporto tra la somma degli scarti al quadrato (ponderati con le rispettive frequenze assolute) ed il collettivo N:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{S} (x_i - \mu)^2 n_i}{N}} = \sqrt{\frac{2.158{,}08}{200}} = \sqrt{10{,}79} = 3{,}38$$

VARIANZA:
Per il calcolo della varianza eseguiamo il quadrato dello scarto quadratico medio:

$$\sigma^2 = \frac{\sum_{i=1}^{S} (x_i - \mu)^2 n_i}{N} = \frac{2.158,08}{200} = 10,79$$

DEVIANZA

Dev(x) è il numeratore della varianza:

$$Dev(\mathrm{X}) = \sum_{i=1}^{S} (x_i - \mu)^2 n_i = 2.158,08$$

CURVA NORMALE E ASIMMETRIA

Si consideri l'altezza di un collettivo di studenti che si distribuisce in maniera normale con:

$$\mu = 165 \qquad \sigma = 4$$

[0.1] Calcolare la frequenza percentuale degli studenti con altezza compresa tra 163 e 168
Fr{163 < x < 168}

[0.2] Calcolare la frequenza percentuale degli studenti con altezza compresa tra 160 e 162.
Fr {160 < x < 162}.

[0.3] Calcolare la frequenza percentuale degli studenti con altezza compresa tra 166 e 169:
Fr {166 < x < 169}.

[0.4] Calcolare la frequenza percentuale degli studenti con altezza superiore a 168:
Fr {x > 168}.

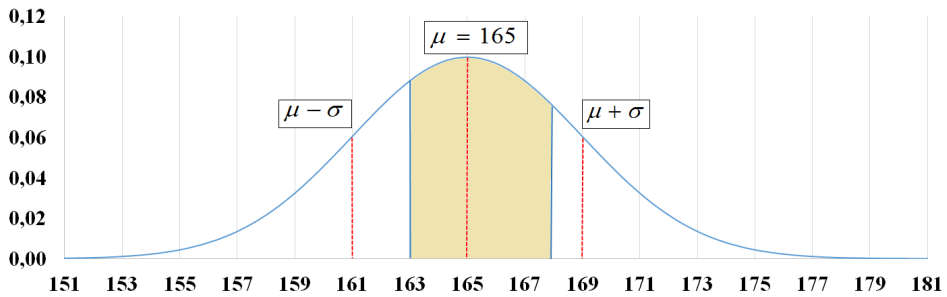
[0.5] Calcolare la frequenza percentuale degli studenti con altezza inferiore a 164:
Fr {x < 164}.

[0.6] Calcolare la probabilità di estrarre dal collettivo uno studente con altezza superiore a 163:
Fr {x > 163}.

[0.7] Calcolare la frequenza percentuale degli studenti con altezza inferiore a 168
Fr {x < 168}.

[0.1] SOLUZIONE

163 < x < 168



$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{163-165}{4} = -\frac{2}{4} = -0{,}5; \qquad z_2 = \frac{x_i - \mu}{\sigma} = \frac{168-165}{4} = \frac{3}{4} = 0{,}75$$

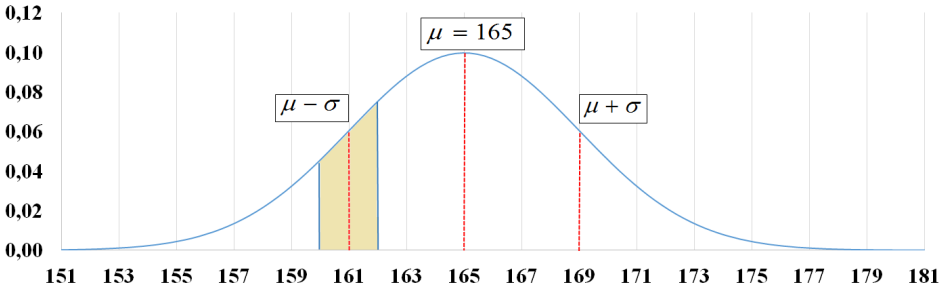
cercando i valori dell'integrale della curva normale standardizzata sulla Tavola in appendice (non avremo valori negativi data la simmetria della distribuzione), avremo:

$$P_{(z_1)} = 0{,}19146; \qquad P_{(z_2)} = 0{,}27337; \qquad P_{(z_1)} + P_{(z_2)} = 0{,}19146 + 0{,}27337 = 0{,}46483$$

Moltiplicando per 100 il valore individuato otteniamo la percentuale di osservazioni comprese nell'intervallo tra 163 e 168 che è il 46%

[0.2] SOLUZIONE

160 < x < 162



$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{160-165}{4} = -\frac{5}{4} = -1{,}25; \qquad z_2 = \frac{x_i - \mu}{\sigma} = \frac{162-165}{4} = -\frac{3}{4} = -0{,}75$$

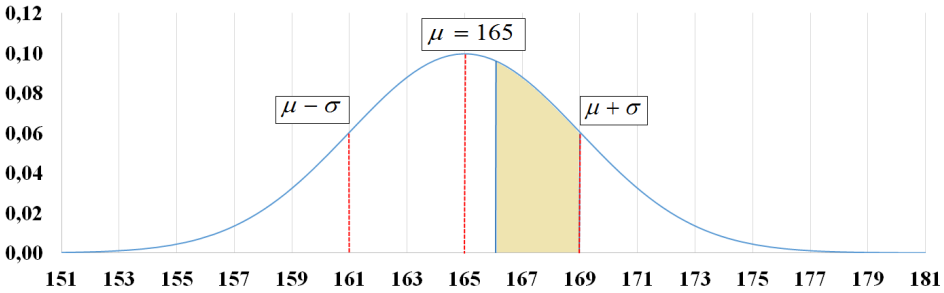
cercando sulla Tavola in appendice i valori dell'integrale della curva normale standardizzata (non si troveranno valori negativi data la simmetria della distribuzione), avremo:

$$P_{(z_1)} = 0{,}39435; \qquad P_{(z_2)} = 0{,}27337; \qquad P_{(z_1)} - P_{(z_2)} = 0{,}39435 - 0{,}27337 = 0{,}12098$$

Il valore ottenuto si può moltiplicare per 100 per conoscere la percentuale di osservazioni comprese nell'intervallo tra 160 e 162 che rappresentano, quindi, il 12% della distribuzione

[0.3] SOLUZIONE

166 < x < 169



$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{166-165}{4} = \frac{1}{4} = 0{,}25; \qquad z_2 = \frac{x_i - \mu}{\sigma} = \frac{169-165}{4} = \frac{4}{4} = 1$$

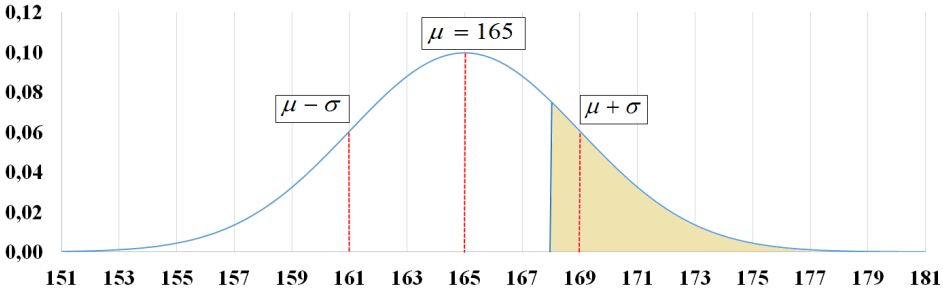
cercando sulla Tavola in appendice i valori dell'integrale della curva normale standardizzata (non si troveranno valori negativi data la simmetria della distribuzione), avremo:

$$P_{(z_1)} = 0{,}09871; \qquad P_{(z_2)} = 0{,}34134; \qquad P_{(z_2)} - P_{(z_1)} = 0{,}34134 - 0{,}09871 = 0{,}24253$$

Il valore ottenuto si può moltiplicare per 100 per conoscere la percentuale di osservazioni comprese nell'intervallo tra 166 e 169 che rappresentano, quindi, il 24% della distribuzione.

[0.4] SOLUZIONE

x > 168



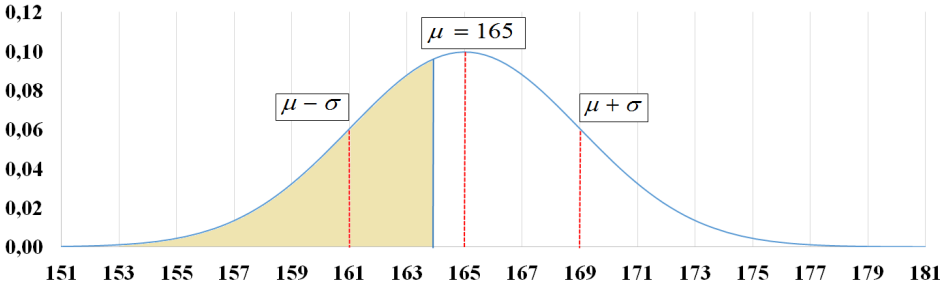
$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{168 - 165}{4} = \frac{3}{4} = 0{,}75;$$

$$P_{(z_1)} = 0{,}27337; \qquad 0{,}5 - P_{(z_1)} = 0{,}5 - 0{,}27337 = 0{,}22563$$

La percentuale delle osservazioni maggiori di 168 corrisponde al 22%.

[0.5] SOLUZIONE

x < 164



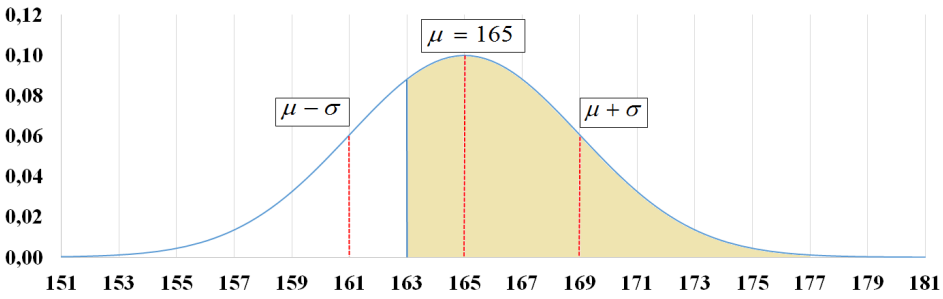
$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{164 - 165}{4} = -\frac{1}{4} = -0{,}25;$$

$$P_{(z_1)} = 0{,}09871; \qquad 0{,}5 - P_{(z_1)} = 0{,}5 - 0{,}09871 = 0{,}40129$$

La percentuale delle osservazioni minori di 164 corrisponde al 40%.

[0.6] SOLUZIONE

x >163



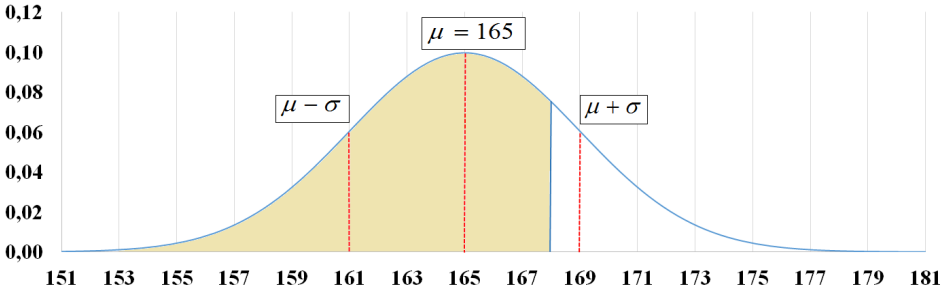
$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{163-165}{4} = -\frac{2}{4} = -0{,}50;$$

$$P_{(z_1)} = 0{,}19146; \qquad 0{,}5 + P_{(z_1)} = 0{,}5 + 0{,}19146 = 0{,}69146$$

La percentuale delle osservazioni maggiori di 163 corrisponde al 69%.

[0.7] SOLUZIONE

x<168



$$z_1 = \frac{x_i - \mu}{\sigma} = \frac{168-165}{4} = \frac{3}{4} = 0{,}75;$$

$$P_{(z_1)} = 0{,}27337; \qquad 0{,}5 + P_{(z_1)} = 0{,}5 + 0{,}27337 = 0{,}0{,}77337$$

La percentuale delle osservazioni minori di 168 corrisponde al 77%.

] SOLUZIONE

x_i	n_i	$x_i\, n_i$	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot n_i$	$(x_i - \bar{x})^3$	$(x_i - \bar{x})^3 \cdot n_i$	$(x_i - \bar{x})^4$	$(x_i - \bar{x})^4 \cdot n_i$
23	3	69	-2,4	5,76	17,28	-13,82	-41,47	33,18	99,53
24	6	144	-1,4	1,96	11,76	-2,74	-16,46	3,84	23,05
25	8	200	-0,4	0,16	1,28	-0,06	-0,51	0,03	0,20
26	5	130	0,6	0,36	1,8	0,22	1,08	0,13	0,65
27	5	135	1,6	2,56	12,8	4,10	20,48	6,55	32,77
28	3	84	2,6	6,76	20,28	17,58	52,73	45,70	137,09
	30	762			65,2		15,85		293,29

Per calcolare l'*indice di asimmetria sk* procediamo utilizzando la formula:

$$sk = \frac{\mu - Mo}{\sigma} = \frac{25{,}4 - 25}{1{,}5} = 0{,}27$$

Calcoliamo *media*, *moda* e *scarto quadratico medio*:

$$\bar{x} = \frac{762}{30} = 25{,}4$$

La moda nella distribuzione è 25

bisogna calcolare sigma $\sigma = \sqrt{\frac{\sum_{i=1}^{S}(\chi_i - \mu)^2 n_i}{N}} = \sqrt{\frac{65{,}2}{30}} = 1{,}5$

Per calcolare il *coefficiente di asimmetria*

$$\gamma_1 = \frac{\sum_{i=1}^{S}(\chi_i - \mu)^3 n_i}{N\sigma^3} = \frac{15{,}85}{96{,}1} = 0{,}16 .$$

e il *coefficiente di eccesso di curtosi*

$$\gamma_2 = \frac{\sum_{i=1}^{S}(\chi_i - \mu)^4 n_i}{N\sigma^4} - 3 = \frac{293{,}3_i}{141{,}7} - 3 = -0{,}93$$