

CAPITOLO 5

- **Pag. 57:** nell'esempio 4.2 è corretta la linea dei numeri del passo 2 mentre quella del passo 3 è errata e va corretta come segue.

La linea ordinata corretta nel Passo 3 è la seguente:

0 6 10 16 22 25 33 41 44 50

Quindi, poiché la posizione della mediana è 5,5, il valore della mediana cadrà tra i valori 22 e 25.

$$\text{Dunque, } Me = \frac{22+25}{2} = 23,5$$

La mediana è uguale a 23,5.

- **pagg. 71-72:** nell'esempio 5.1 inserire il valore 4 con frequenza zero nella tabella.

Tablelle errate:

Passo 2 Calcolare le frequenze cumulate a partire dalle frequenze osservate.

Livello socio-economico	f_i	f_{cum}
1	2	2
2	2	4
3	4	8

yout1 30/01/17 14:28 Pagina 72

Gli indici di posizione

Livello socio-economico	f_i	f_{cum}
5	6	14
6	2	16
7	8	24
8	4	28
9	4	32
10	2	34
11	4	38
12	2	40

Livello socio-economico	f_i
1	2
2	2
3	4
5	6
6	2
7	8
8	4
9	4
10	2
11	4
12	2

Livello socio-economico	f_i	f_{cum}
1	2	2
2	2	4
3	4	8
5	6	14 ← Q_1
6	2	16
7	8	24 ← Q_2
8	4	28
9	4	32 ← Q_3
10	2	34
11	4	38
12	2	40

Tablelle corrette:

Livello socio-economico	f_i
1	2
2	2
3	4
4	0
5	6
6	2
7	8
8	4
9	4
10	2
11	4
12	2

Livello socio-economico	f_i	f_{cum}
1	2	2
2	2	4
3	4	8
4	0	8
5	6	14
6	2	16
7	8	24
8	4	28
9	4	32
10	2	34
11	4	38
12	2	40

Livello socio-economico	f_i	f_{cum}	
1	2	2	
2	2	4	
3	4	8	
4	0	8	
5	6	14 ← Q_1	
6	2	16	
7	8	24 ← Q_2	
8	4	28	
9	4	32 ← Q_3	
10	2	34	
11	4	38	
12	2	40	

- **pag. 96:** la formula che compare per dati raggruppati, al denominatore dovrebbe essere N e non $N+1$. Questa cosa va modificata anche nell'eserciziario relativo al capitolo 7; dunque:

formula errata: $RP(X) = \frac{PosP_X \cdot quantile}{N+1}$

formula corretta: $RP(X) = \frac{PosP_X \cdot quantile}{N}$

CAPITOLO 6

- **Pag. 103**, esempio 6.3 – passo 2: sostituire la parola “presenza” con “presenze”;
- **Pag. 107**, paragrafo 6.7: aggiungere nel secondo rigo il verbo “calcolare” a seguire la parola “possiamo”, rendendo la frase: “Lo scostamento semplice medio (SSM) è un indicatore di dispersione o variabilità che possiamo calcolare a partire da scala ad intervalli equivalenti...”;
- **Pag. 107**, paragrafo 6.7, quinto rigo: cancellare la parola “sua”, rendendo la frase: “Prima di procedere a descrivere la modalità di calcolo...”.
- **Pag. 111**, Esempio 6.7: sostituire in tabella, in corrispondenza del paziente F, 20 a 30.

Tabella Errata:

Paziente	X_i
A	20
B	15
C	9
D	11
E	32
F	30
G	23
H	16
I	9
L	24
M	35
N	28
O	20
P	9
Q	9

Tabella Corretta:

Paziente	X_i
A	20
B	15
C	9
D	11
E	32
F	20
G	23
H	16
I	9
L	24
M	35
N	28
O	20
P	9
Q	9

CAPITOLO 7

- **Pag. 141**, esempio 7.2, Caso 1, passo 3: il valore dell'area sottesa relativa al punto z 1,44 è 0,4251.

Dunque, "ciò vuol dire che il 42,51% della popolazione ottiene un punteggio che cade all'interno dell'intervallo delimitato dai punti z 0 e 1,44".

- **Pag. 149**, esempio 7.6, passo 4: nello svolgimento della formula sostituire il numero 14 con il numero 6, come segue:

$$PosP_x = 8 + \frac{16-11,5}{6} \times 6 = 8 + 0,75 \times 6 = 12,5$$

CAPITOLO 8

- **pag.167**: l'affermazione: "il secondo caso è quello in cui entrambe le variabili da mettere in correlazione sono misurate su scala ordinale, ma i dati non sono ordinati in graduatoria." è incongrua con l'esempio nel testo (non nel riquadro!) che segue. Infatti, i dati sono già in graduatoria (poiché su scala ordinale) ma vanno solo ordinati in senso crescente. Nell'esempio, invece, si fa riferimento a punteggi su scala quantitativa e relativa trasformazione in scala ordinale, che sembra più inerente al terzo caso, ovvero "una variabile ordinale e l'altra continuativa". Dunque, l'esempio nel testo (non nel riquadro!) nel testo andrebbe spostato più sotto, quando sia affronta il caso tre.

Dunque:

A pag. 169 dovrebbe essere "Il secondo caso è quello in cui entrambe le variabili da mettere in correlazione sono misurate su scala ordinale, ma i dati non sono ordinati in graduatoria. In questo caso, bisogna ordinare la graduatoria, costruendo una tabella dove nella, nella prima colonna sono riportati tutti i soggetti che fanno parte del campione e nella seconda e terza colonna le loro posizioni, rispettivamente alla variabile X e alla variabile Y. L'Esempio 8.4 mostra un'applicazione del procedimento di calcolo del coefficiente di correlazione r_s di Spearman quando i dati devono essere ordinati in graduatoria."

A pag. 171 va aggiunto il pezzo sostituito a pagina 169 (da "A tal punto..." fino a "dovrebbero essere ordinati in graduatoria." prima dell'affermazione "L'Esempio 8.5 chiarirà in maniera pratica questo terzo caso."

- **pag. 178:** nella formula 8.7, in didascalia, si legge "dove (per X = variabile continua e Y = variabile dicotomica che assume i valori a o b)" mentre nell'esercizio relativo (Esempio 8.8, pag 179) le variabili sono invertite (la X è dicotomica mentre Y è continua).

Eliminare, sotto la formula, la frase "dove (per X = variabile continua e Y = variabile dicotomica che assume i valori a o b)".

- **pag. 186:** nella formula 8.12, al denominatore, manca la radice quadrata (infatti, poco prima di afferma Sostituendo al denominatore la formula della deviazione standard, si avrà:";

formula errata:

$$r_{xy} = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\left(\frac{\sum X^2}{N} - \bar{X}^2\right) \left(\frac{\sum Y^2}{N} - \bar{Y}^2\right)}$$

formula corretta:

$$r_{XY} = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\left(\frac{\sum X_i^2}{N} - \bar{X}^2\right) \cdot \left(\frac{\sum Y_i^2}{N} - \bar{Y}^2\right)}}$$

- **pag. 186,** manca "i" al pedice ad X e Y del denominatore:

formula errata:

$$r_{xy} = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum \left(\frac{X}{N} - \bar{X}\right)^2 \sum \left(\frac{Y}{N} - \bar{Y}\right)^2}}$$

formula corretta: $r_{XY} = \frac{\sum[(X_i - \bar{X}) \cdot (Y_i - \bar{Y})]}{\sqrt{\sum\left(\frac{X_i - \bar{X}}{N}\right)^2 \cdot \sum\left(\frac{Y_i - \bar{Y}}{N}\right)^2}}$

CAPITOLO 09

- **pag. 191**: si afferma : "A questo punto, possiamo facilmente calcolare la proporzione di "varianza residua", ossia di varianza della VD che non viene spiegata in base alla relazione lineare con la VD, che è uguale a ". Si ripete due volte VD. dovrebbe essere : "A questo punto, possiamo facilmente calcolare la proporzione di "varianza residua", ossia di varianza della VD che non viene spiegata in base alla relazione lineare con la VI..."

Frase errata: "A questo punto, possiamo facilmente calcolare la proporzione di "varianza residua", ossia di varianza della VD che non viene spiegata in base alla relazione lineare con la VD, che è uguale a:.."

Frase corretta: "A questo punto, possiamo facilmente calcolare la proporzione di "varianza residua", ossia di varianza della VD che non viene spiegata in base alla relazione lineare con la VI, che è uguale a:.."

- **pag. 195**: Il coefficiente di regressione (o coefficiente angolare) può assumere tutti i valori. È il coefficiente di correlazione che varia da -1,00 a +1,00.

CAPITOLO 11

- **pag. 233**: Undicesimo rigo del secondo capoverso: cancellare la congiunzione "se" che segue il punto 2, quindi rendere la frase: "...2) il test di cui si valuta la validità presenta qualche vantaggio incrementale rispetto al test ctiterio..."

Nella sezione: **Ricorda che...**

- **Pag. 128** → la tabella deve essere la seguente

Segno n_1	Segno n_2	Segno $n_1 \times n_2$ ($n_1 : n_2$)
+	+	+
+	-	-
-	+	-
-	-	+

- **Pag. 130**

La formula sbagliata è $s^2 = \frac{\sum f_i \cdot X_c^2}{N} - (\bar{X})^2$ che va dunque tolta.

La formula corretta, e che va inserita è $s^2 = \frac{\sum f_i \cdot X_i^2}{N} - (\bar{X})^2$

Nella sezione: **Esercizi con soluzioni:**

- **Pag. E4. L'esercizio 2 va ristampato in questo modo.**

In uno studio effettuato su un campione di 120 soggetti residenti nel comune di Torino vengono indagati gli interessi culturali degli intervistati. Viene chiesto loro quale tra le seguenti attività culturali hanno svolto con maggior frequenza nell'ultimo anno: leggere un libro, andare al cinema, andare ad una mostra, visitare un sito archeologico, andare ad un concerto, assistere ad uno spettacolo teatrale, oppure nessuna. Riassumiamo i risultati nella tabella sottostante:

Attività culturale	f
libro	32
Cinema	37
mostra	9
sito archeologico	6
concerto	13
spettacolo teatrale	5
nessuna	18

1. Individuare la scala di misura della variabile;
2. Costruire la distribuzione di frequenza percentuale;
3. Rappresentare la distribuzione di frequenza percentuale mediante un grafico a barre;
4. Rappresentare la distribuzione di frequenza percentuale mediante un grafico a torta.

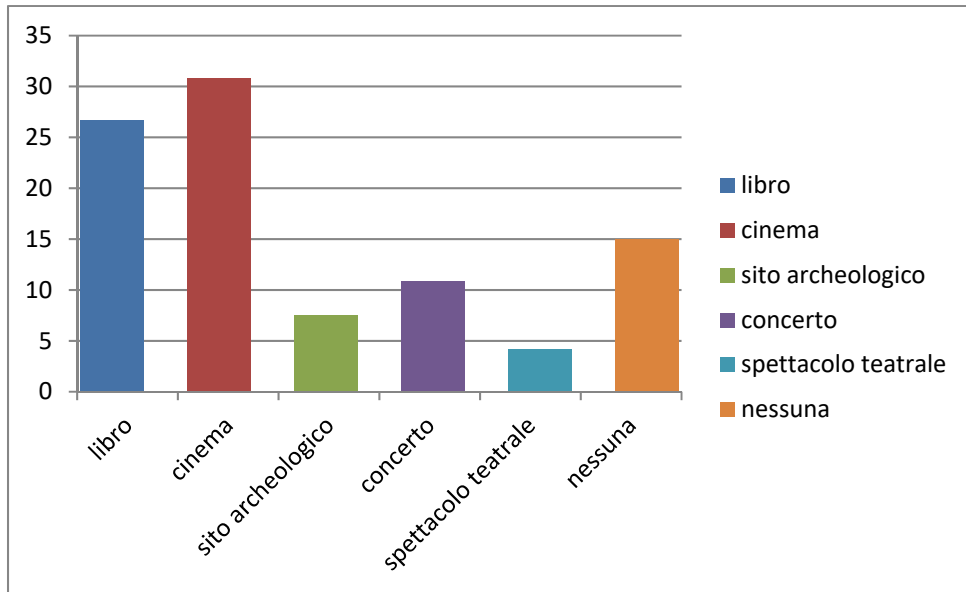
Soluzione esercizio 2:

1. La variabile attività culturale svolta con maggior frequenza nell'ultimo anno è una variabile nominale, poiché la scelta del soggetto ricade su una categoria, e nel caso specifico non esistono relazioni di ordine tra le diverse categorie.
2. Costruiamo la distribuzione di frequenza percentuale:

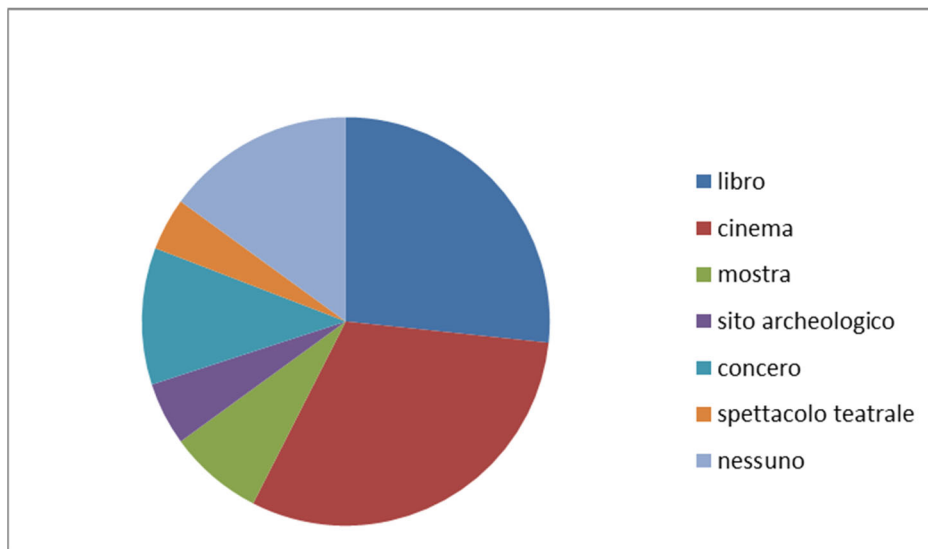
Attività culturale	Ff1	Ff1%
Libro	32	(32/120)×100= 26,66 %
Cinema	37	(37/120)×100= 30,83 %
Mostra	9	(9/12)×100= 7,50 %
Sito archeologico	6	(6/12)×100= 5,00 %
Concerto	13	(13/120)×100= 10,83 %

Spettacolo teatrale	5	$(5/120) \times 100 =$ 4,16 %
Nessuna	18	$(18/120) \times 100 =$ 15,00 %
	120	100 %

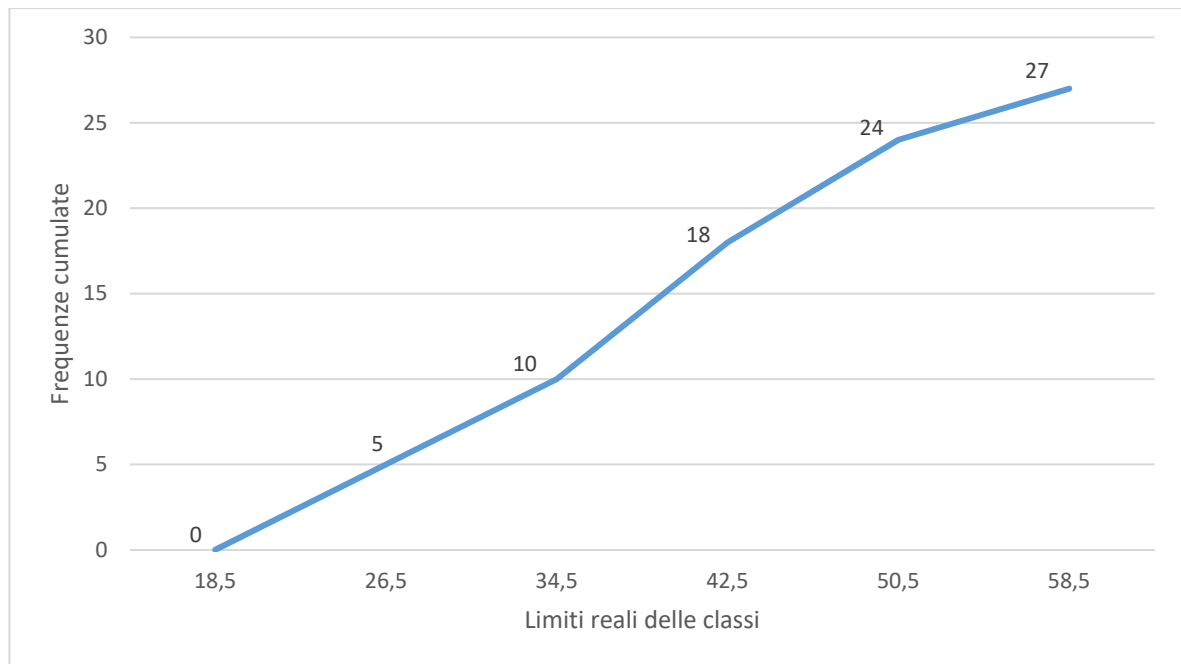
3. Rappresentiamo graficamente la distribuzione di frequenza percentuale mediante grafico a barre:



4. Rappresentiamo graficamente la distribuzione di frequenza percentuale mediante grafico a torta:



- **Pag. E11:** Esercizio 5, Soluzione – punto 3: Correggere nella tabella in fondo alla pagina la scritta “classie” nella prima e nella seconda colonna, sostituendo la parola con “classi”.
- **Pag. E12-** esercizio 5, passo 4: sostituire il grafico presente nel libro, con il seguente:



- **Pag. E18:** In corrispondenza della formula della varianza per dati grezzi, oltre a quella già presente, aggiungere la formula: $s^2 = \frac{\sum(x-\bar{x})^2}{N}$
- **Pag. E25:** il quinto passaggio non è “Calcolare il settantesimo percentile (P₇₀)” ma “Calcolare il cinquantesimo percentile (P₅₀)”

Frase errata: “Calcolare il settantesimo percentile (P₇₀)”

Frase corretta: “Calcolare il cinquantesimo percentile (P₅₀)”

- **pag. E44,** aggiungere la “**Formula di calcolo del Rango Percentile per dati raggruppati**”

$$RP(X) = \frac{PosP_x \cdot 100}{N}$$

- **pag. E50,** soluzione Esercizio 4:

Svolgimento errato:

$$RP(X) = \frac{PosP_X \times 100}{N + 1}$$

$$RP(11) = \frac{8,60 \cdot 100}{30 + 1} = \frac{860}{31} = 27,74$$

Il Rango Percentile del punteggio $X = 11$ è pari a 27,74. Ciò significa che il 27,74% della popolazione ottiene un punteggio uguale o inferiore a 11.

Svolgimento corretto

$$RP(X) = \frac{PosP_X \cdot 100}{N}$$

$$RP(11) = \frac{8,60 \cdot 100}{30} = \frac{860}{30} = 28,66$$

Il Rango percentile del punteggio $X = 11$ è pari a 28,66. Ciò significa che il 28,66 della popolazione ottiene un punteggio uguale o inferiore ad 11.

Pagina E53

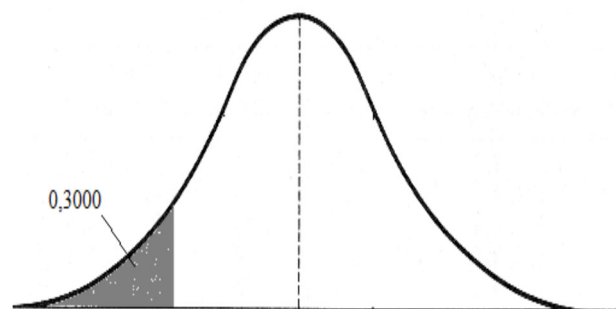
ESERCIZIO 8

Utilizzando i dati dell'esercizio 7, individuare il punteggio X_i che corrisponde al terzo decile (D_3), ovvero a quel punteggio al di sotto del quale cade il 30% delle osservazioni.

Soluzione:

Il terzo decile (D_3) è quel valore della distribuzione al di sotto del quale si trova il 30% dei casi della distribuzione.

Attraverso la rappresentazione grafica della curva normale standardizzata, possiamo notare facilmente come il 30% dei soggetti corrisponda all'area 0,3000 (30/100) nella coda della metà della curva a sinistra della media, evidenziata in grigio nella figura seguente:



Per poter conoscere l'area evidenziata in grigio, sfruttiamo le proprietà della distribuzione normale standardizzata: sapendo che l'intera metà sinistra è uguale a 0,5, sottraiamo a tale valore l'area che conosciamo (0,3000). Dunque:

$$0,5 - 0,3000 = 0,2000$$

Consultando la tavola della distribuzione normale standardizzata, al suo interno, in corrispondenza di un'area di 0,2000 (e, precisamente, di 0,1950) individueremo un punto $z = 0,52$, al quale anticiperemo il segno negativo poiché ci troviamo nella metà curva al di sotto della media, che contiene punti z negativi.

A questo punto, possiamo calcolare il punteggio x che delimita la coda del 30% dei soggetti meno estroversi, con la nota formula inversa:

$$D_3 = \mu + z_i \cdot \sigma = 50 + (-0,52 \cdot 5) = 50 - 2.60 = 47,4$$

Possiamo, quindi, concludere che il punteggio $X=47,4$ è il punteggio al di sotto del quale si trova il 30% dei casi. In altri termini, il 30% dei soggetti meno creativi della popolazione considerata hanno ottenuto un punteggio uguale o inferiore a 47,4.