

STATISTICA (I MODULO - STATISTICA DESCRITTIVA)
Soluzione esercitazione 5

Esercizio A.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
x_i	245	245	247	247	238	245	248	241	248	246	245	249	244	243	246	3677
$(x_i - \mu)^2$	0,02	0,02	3,48	3,48	50,88	0,02	8,22	17,08	8,22	0,75	0,02	14,95	1,28	4,55	0,75	113,73
$(x_i - \mu)^3$	0,00	0,00	6,50	6,50	-362,98	0,00	23,56	-70,62	23,56	0,65	0,00	57,81	-1,46	-9,71	0,65	-325,53
y_i	238	241	243	244	245	245	245	245	246	246	247	247	248	248	249	3677
F_i	0,067	0,133	0,200	0,267	0,333	0,400	0,467	0,533	0,600	0,667	0,733	0,800	0,867	0,933	1,000	

- Il peso medio del lotto è pari a $\mu = \sum_{i=1}^{15} x_i/15 = 3677/15 = 245,13$, mentre la deviazione standard è pari a $\sigma = \sqrt{\sum_{i=1}^{15} (x_i - \mu)^2/15} = \sqrt{113,73/15} = 2,75$
- L'indice di asimmetria richiesto è il seguente

$$\alpha_1 = \frac{\sum_{i=1}^{15} (x_i - \mu)^3/15}{\sigma^3} = \frac{-325,53/15}{2,75^3} = -1,04$$

il quale indica un'asimmetria negativa.

- Le quantità necessarie per disegnare il diagramma a scatola sono le seguenti:

$$\begin{aligned} y_1 &= 238 \\ q_1 &= 244 \\ m &= 245 \\ q_3 &= 247 \\ y_{15} &= 249 \end{aligned}$$

Il corrispondente grafico è riportato in Figura 1, da cui si può vedere come la distribuzione nel suo complesso presenti un'asimmetria negativa.

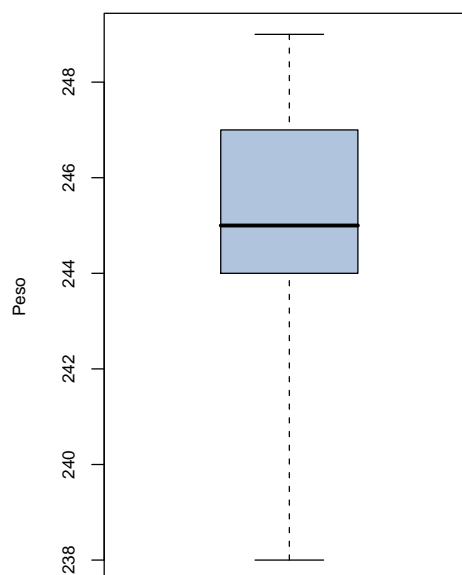


Figura 1

Esercizio B. Le quantità necessarie per rispondere ai quesiti, in parte calcolate al punto C dell'esercitazione 4, sono riportate nella tabella seguente:

	Occupati	In cerca di occupazione
min	15,000	15,000
q_1	32,883	24,938
m	40,930	32,012
q_3	49,795	41,428
max	69,000	69,000

Si noti che per il minimo ed il massimo delle due distribuzioni sono stati assegnati i valori corrispondenti, rispettivamente, all'estremo inferiore della prima classe e all'estremo superiore dell'ultima classe.

- L'indice di asimmetria per la distribuzione dell'età tra gli occupati è pari a

$$\alpha_2 = \frac{(49,795 - 40,930) - (40,930 - 32,883)}{(49,795 - 40,930) + (40,930 - 32,883)} = 0,0484$$

mentre per quella di coloro in cerca di occupazione è pari a

$$\alpha_2 = \frac{(41,428 - 32,012) - (32,012 - 24,938)}{(41,428 - 32,012) + (32,012 - 24,938)} = 0,1420$$

In entrambe è presente un lieve grado di asimmetria positiva, maggiore tra coloro che sono in cerca di occupazione rispetto agli occupati.

2. Il grafico in Figura 2 mostra i diagrammi a scatola per le due distribuzioni. L'asimmetria è graficamente impercettibile per gli occupati, mentre risulta più evidente per coloro in cerca di occupazione, soprattutto in corrispondenza delle code inferiore e superiore della distribuzione.

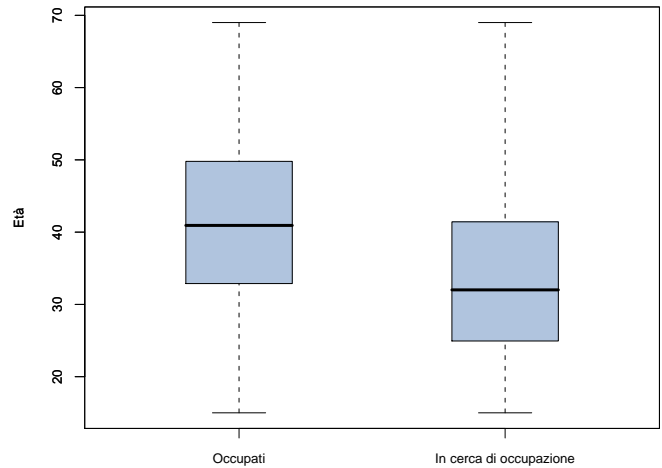


Figura 2

Esercizio C.

1. La seguente tabella riporta la distribuzione congiunta dei due caratteri. Le distribuzioni marginali compaiono ai lati della tabella in corrispondenza dei totali di riga e colonna:

	Agricoltura	Industria	Terziario	In cerca di occupazione	Totale
Nord-Ovest	157	2361	4356	270	7144
Nord-Est	189	1812	3047	162	5210
Centro	122	1271	3392	267	5052
Mezzogiorno	456	1560	4500	808	7324
Totale	924	7004	15295	1507	24730

Quindi, le distribuzioni marginali sono le seguenti:

Zona geografica	n_{i0}	Condizione professionale	n_{0j}
Nord-Ovest	7144	Agricoltura	924
Nord-Est	5210	Industria	7004
Centro	5052	Terziario	15295
Mezzogiorno	7324	In cerca di occupazione	1507
Totale	24730	Totale	24730

2. Le distribuzioni condizionate percentuali rispetto all'area geografica di residenza sono le seguenti:

	Agricoltura	Industria	Terziario	In cerca di occupazione	Totale
Nord-Ovest	2,20	33,05	60,97	3,78	100
Nord-Est	3,63	34,78	58,48	3,11	100
Centro	2,41	25,16	67,14	5,29	100
Mezzogiorno	6,23	21,30	61,44	11,03	100

3. Le distribuzioni condizionate percentuali del punto precedente risultano lievemente differenti, in particolare quella del Mezzogiorno e del Centro rispetto alle zone Nord del paese. Qualora i due caratteri fossero stati indipendenti tali distribuzioni sarebbero risultate uguali. Per tale motivo ci aspettiamo la presenza di un certo grado di dipendenza.

4. La tabella di indipendenza in cui le frequenze congiunte sono calcolate come $n_{ij}^* = n_{i0}n_{0j}/N$ è la seguente:

	Agricoltura	Industria	Terziario	In cerca di occupazione	Totale
Nord-Ovest	266,925	2023,315	4418,418	435,342	7144
Nord-Est	194,663	1475,570	3222,279	317,488	5210
Centro	188,761	1430,821	3124,559	307,859	5052
Mezzogiorno	273,651	2074,294	4529,744	446,311	7324
Totale	924	7004	15295	1507	24730

L'indice di dipendenza χ^2 può essere calcolato come segue:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = \frac{(157 - 266,925)^2}{266,925} + \frac{(2361 - 2023,315)^2}{2023,315} + \dots + \frac{(808 - 446,3109)^2}{446,311} = 939,969$$

L'indice relativo di dipendenza è dato da

$$C = \sqrt{\frac{\chi^2}{N \min(4 - 1, 4 - 1)}} = \sqrt{\frac{939,969}{24730 \times 3}} = 0,1126$$

che mostra la presenza di un certo grado di dipendenza, seppure non molto elevata.

Esercizio D.

1. La seguente tabella riporta la distribuzione congiunta dei due caratteri, mentre le distribuzioni marginali compaiono ai lati della tabella:

	Fino a 1	1-2	2-3	3-5	5-10	Oltre 10	Totale
Volontario	0	19	40	148	181	647	1035
Colposo	2190	184	8	5	1	0	2388
Totale	2190	203	48	153	182	647	3423

2. Le distribuzioni condizionate percentuali per tipo di reato sono le seguenti:

	Fino a 1	1-2	2-3	3-5	5-10	Oltre 10	Totale
Volontario	0,00	1,84	3,86	14,30	17,49	62,51	100
Colposo	91,71	7,71	0,34	0,21	0,04	0,00	100

Come si può vedere le due distribuzioni risultano molto differenti tra di loro. In particolare, un'elevata percentuale (circa il 91%) dei reati colposi ha visto pene fino ad 1 anno, mentre per i reati volontari le pene sono molto più elevate, con circa il 62% di essi a cui sono stati inflitti più di 10 anni di detenzione.

3. La tabella di indipendenza è la seguente:

	Fino a 1	1-2	2-3	3-5	5-10	Oltre 10	Totale
Volontario	662,182	61,380	14,514	46,262	55,031	195,631	1035
Colposo	1527,818	141,620	33,486	106,738	126,969	451,369	2388
Totale	2190	203	48	153	182	647	3423

L'indice di dipendenza χ^2 può essere calcolato come segue:

$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - n_{ij}^*)^2}{n_{ij}^*} = \frac{(0 - 662,182)^2}{662,182} + \frac{(19 - 61,380)^2}{61,380} + \dots + \frac{(0 - 451,369)^2}{451,369} = 3282,11$$

L'indice relativo di dipendenza è dato da

$$C = \sqrt{\frac{\chi^2}{N \min(2 - 1, 6 - 1)}} = \sqrt{\frac{3282,11}{3423 \times 1}} = 0,9792$$

che mostra la presenza di un elevato grado di dipendenza tra il tipo di reato commesso e la pena detentiva inflitta.