

Olimpiadi italiane di statistica

Anno 2020

Prova individuale

Classi III e IV

Esercizi con soluzioni

QUESITO 1

La tabella rappresenta il numero di matrimoni nel 2017 per le prime tre classi di età dello sposo e della sposa (fonte: dati.istat.it):

Età al matrimonio dello sposo	Età al matrimonio della sposa		
	16-19 anni	20-24 anni	25-29 anni
16-19 anni	79	62	17
20-24 anni	618	2961	1144
25-29 anni	340	6593	20698

Per ogni classe di età si consideri il valore centrale della classe, ossia:

- ✓ 17,5 anni per la classe 16-19 anni
- ✓ 22 anni per la classe 20-24
- ✓ 27 anni per la classe 25-29

Quale di queste affermazioni è vera:

- **La covarianza fra le età dello sposo e della sposa è positiva e pari a 2,11**
- La covarianza fra le età dello sposo e della sposa è negativa e pari a -2,11
- La covarianza fra le età dello sposo e della sposa è pari a zero
- Non può essere che la covarianza sia pari a 2,11 o -2,11 perché la covarianza assume valori compresi fra -1 e 1
- Non so

Soluzione:

Anche solo osservando la tabella si può notare come la covarianza tra i due caratteri, cioè la misura di come essi varino assieme, ovvero della loro dipendenza, sia positiva (com'è intuitivo che sia, date la due variabili in esame). Infatti, possiamo notare, per riga in particolare, ovvero per età dello sposo, che ogni volta la casella che presenta i numeri più alti sia proprio quella corrispondente alla medesima classe di età della sposa. Ovviamente possiamo anche risolvere l'esercizio calcolando la suddetta covarianza, attraverso la formula:

$$\sigma_{x,y} = \frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Quindi dovremmo dapprima calcolare la media di x e quella di y (utilizzando, come suggerito, i valori centrali delle classi di età), quindi sommare i prodotti delle differenze dei vari x_i dalla media di x e dei vari y_i dalla media di y e dividere il tutto per N . Sono calcoli abbastanza lunghi ed onerosi, cui si può ovviare con la semplice attenta osservazione della tabella e dei numeri in essa contenuti, osservazione che ci porta ad accettare la prima risposta come quella corretta; tutte le altre sono false.

QUESITO 2

La Rilevazione sulle Forze di Lavoro condotta dall'Istat fornisce, tra gli altri, dati sui giovani che abbandonano precocemente la scuola, ossia ragazzi di età compresa tra i 18 e i 24 anni che hanno abbandonato gli studi con al massimo un titolo di studio secondario inferiore, che non hanno concluso un corso di formazione professionale e che non frequentano corsi scolastici né svolgono attività formative. La tabella seguente fornisce le frequenze percentuali dei giovani che abbandonano precocemente la scuola rispetto al sesso in ognuna delle quattro regioni del Nord-ovest e il numero di maschi residenti nella stessa classe di età:

Regione	Percentuale di giovani che abbandonano la scuola		Numero di maschi residenti di età compresa fra 18 e 24 anni
	Sesso		
	Maschi	Femmine	
Piemonte	16,2	9,0	145654
Valle d'Aosta / Vallée d'Aoste	18,9	13,4	4342
Liguria	15,7	11,5	49669
Lombardia	15,4	10,2	348896

Si determini la percentuale di coloro che abbandonano prematuramente gli studi fra i maschi di età compresa fra i 18 e i 24 anni residenti nelle 4 regioni del Nord-ovest.

Quale frequenza percentuale è corretta?

- F=16,6%
- F=66,2%
- F=15,7%**
- F=14,3%
- Non so

Soluzione:

Per calcolare la percentuale richiesta dobbiamo dapprima calcolare il numero di ragazzi che abbandonano la scuola per ognuna delle 4 regioni. Si ha per il Piemonte, ad esempio:

Ragazzi che abbandonano la scuola in Piemonte = $145654 \cdot 16,2 / 100 = 23596$

Ovvero si deve calcolare il 16,2% del totale dei ragazzi della classe di età in esame per il Piemonte. Eseguendo calcoli simili si ottiene per la Valle d'Aosta 821 ragazzi, per la Liguria 7798 e per la Lombardia 53730. Sommando tali cifre si ottiene il totale di 85945 ragazzi. Ora si deve effettuare la somma di tutti i ragazzi nella classe di età 18-24, ottenendo 548561. A questo punto si può calcolare la percentuale richiesta:

Percentuale di ragazzi del Nord-Ovest che abbandonano gli studi = $85945 / 548561 \cdot 100 = 15,7$

La terza risposta è, dunque, quella corretta. Sarebbe, ovviamente, scorretto eseguire una media semplice delle quattro percentuali per regione per rispondere al quesito, dato che ogni regione ha numerosità anche molto diversa.

QUESITO 3

L'Indagine sugli aspetti della vita quotidiana condotta dall'Istat rileva che nel 2017 fra i maschi di età fra 6 e 24 anni sono il 71,3% coloro che usano il pc. In più, fra coloro che usano il pc, la distribuzione percentuale della frequenza di utilizzo è rappresentata nella tabella seguente:

Frequenza di utilizzo del computer	Percentuale
Qualche volta all'anno	3,2
Qualche volta al mese	8,6
Una o più volte alla settimana	43,0
Tutti i giorni	45,2

Si indichi il valore percentuale sull'intero collettivo dei maschi di età compresa fra 6 e 24 anni di coloro che utilizzano il computer con una frequenza pari a qualche volta al mese o meno spesso:

- 3,2%
- 11,8%
- 2,3%
- 8,4%**
- Non so

Soluzione:

Si tenga conto che le percentuali indicate nel testo hanno significati diversi. Il valore 71,3% è una percentuale che fa riferimento all'intero collettivo dei maschi. Le percentuali indicate in tabella, invece, fanno riferimento al collettivo dei maschi che usano il pc, che è un sottoinsieme del collettivo precedente (ossia il 71,3% del collettivo dei maschi corrisponde al 100% del collettivo dei maschi che usano il pc).

Per determinare il valore percentuale richiesto dobbiamo quindi eseguire il seguente semplice calcolo:

$$\text{Percentuale di ragazzi che usano il pc una volta al mese o meno spesso} = (3,2 + 8,6) \cdot 71,3 / 100 = 8,4$$

Ovvero dobbiamo sommare le due percentuali di chi utilizza il pc qualche volta l'anno e qualche volta al mese, moltiplicare tale somma per la percentuale di quanti utilizzano il pc sull'intera popolazione e dividere il risultato per cento. In tal modo otteniamo una percentuale riferita all'intera popolazione e non più al solo sottoinsieme di quanti utilizzano il computer, come richiesto dal testo dell'esercizio.

QUESITO 4

L'indagine Istat su viaggi e vacanze ha rilevato che nel 2018 il numero di persone che non sono andate in vacanza per motivi economici segue la seguente distribuzione per età:

Classe di età	Numero di persone (in migliaia)
15-24 anni	1502
25-34 anni	1735
35-44 anni	2134
45-54 anni	2678
55-64 anni	2221
65 anni e più	2889

Si determini la classe di età mediana:

- 25-34 anni
- 35-44 anni
- 45-54 anni**
- 55-64 anni
- Non so

Soluzione:

Per determinare la classe di età mediana della distribuzione costruiamo la tabella con le frequenze assolute cumulate:

Classe di età	Numero di persone (in migliaia)	Frequenze assolute cumulate (in migliaia)
15-24 anni	1502	1502
25-34 anni	1735	3237
35-44 anni	2134	5371
45-54 anni	2678	8049
55-64 anni	2221	10270
65 anni e più	2889	13159

Calcoliamo ora la metà del totale, che è $6579,5 (= 13159/2)$. Possiamo ora affermare che la classe di età mediana della distribuzione è quella 45-54 anni, visto che per la classe di età immediatamente precedente la frequenza cumulata somma a 5371 e per la suddetta classe a 8049, valori che comprendono quello che divide a metà il totale (proprio 6579,5, appena calcolato); quindi con la classe di età 35-44 anni siamo a meno del 50% delle persone e con la classe di età 45-54 a più del 50% delle persone (avremmo potuto calcolare la mediana anche utilizzando le frequenze percentuali cumulate, ma con un aggravio di calcoli non necessario).

QUESITO 5

11 ragazzi di una squadra di calcio mettono a confronto i gol segnati da ognuno di loro nel corso di un torneo:

0; 0; 2; 3; 4; 5; 5; 5; 5; 7; 8

Alla fine del torneo colui che ha segnato 3 gol viene sostituito da un nuovo giocatore che nel corso dello stesso torneo ha segnato un numero di gol pari alla media della squadra. Cosa succede alla media aritmetica, alla mediana e alla varianza?

- I tre parametri rimangono inalterati
- La mediana rimane stabile, la media aritmetica cresce, la varianza diminuisce**
- La mediana rimane stabile, la media aritmetica e la varianza crescono
- Tutti e tre i parametri crescono
- Non so

Soluzione:

Per prima cosa calcoliamo la media aritmetica della distribuzione in esame:

$$\text{Media aritmetica} = (0+0+2+3+4+5+5+5+5+7+8)/11 = 44/11 = 4$$

Quindi il numero di gol segnati da ciascun giocatore, per com'è composta la squadra alla fine del torneo, sarà il seguente:

0; 0; 2; 4; 4; 5; 5; 5; 5; 7; 8

Anche senza eseguire calcoli possiamo notare che, nelle due distribuzioni, la mediana rimane la stessa (uguale a 5, valore che occupa in entrambe le distribuzioni il posto centrale, ovvero il 6° posto); la media aritmetica aumenta, seppure di poco, visto che abbiamo sostituito un giocatore che aveva segnato 3 con uno che ne ha segnati 4; ed infine la varianza diminuisce, seppure di poco, visto che nel calcolarla sostituiamo un valore che ha uno scostamento dalla media di 1 con un valore che ha uno scostamento dalla nuova media vicino allo zero (la nuova media aritmetica della seconda distribuzione sarà di poco superiore a 4, con la sostituzione del giocatore, com'è facile verificare attraverso calcoli elementari). L'affermazione corretta è, quindi, la seconda.

QUESITO 6

Dalla Rilevazione sulle forze di lavoro condotta dall'Istat nel 2018 risulta che il numero di giovani nella classe di età 15-19 anni in base alla condizione professionale (occupato/disoccupato), al sesso (maschio/femmina) e alla condizione di studente (sì/no) segue la seguente tabella (valori in migliaia):

Condizione professionale	Studente	Maschi	Femmine	Totale
Occupati	No	62	27	89
	Sì	10	6	16
	Totale	72	33	105
Disoccupati	No	56	39	95
	Sì	5	3	8
	Totale	61	42	103

Quali di queste affermazioni è vera:

- Fra i maschi, è più facile essere disoccupati che occupati
- Fra gli occupati, è più facile essere femmina che maschio
- I maschi sono ugualmente frequenti sia fra gli studenti occupati che fra gli studenti disoccupati**
- Fra le femmine occupate, è più facile essere studente
- Non so

Soluzione:

Se calcoliamo la percentuale dei maschi tra gli studenti occupati otteniamo:

Percentuale dei maschi tra studenti occupati = $10/16 \cdot 100 = 62,5$

Se calcoliamo la stessa percentuale ma per i maschi studenti disoccupati otteniamo:

Percentuale dei maschi tra studenti disoccupati = $5/8 \cdot 100 = 62,5$

Le due percentuali sono uguali, per cui l'affermazione vera è la terza, com'è facile verificare anche senza eseguire i suddetti facili calcoli ma semplicemente riconoscendo che tra gli studenti il numero di maschi e femmine è esattamente doppio per gli occupati rispetto ai disoccupati, per cui la frequenza dei maschi è la stessa nei due casi. Le altre affermazioni sono tutte false, com'è facile verificare osservando i dati della tabella.

QUESITO 7

Si consideri la variabile “spesa mensile” osservata su tre famiglie nel corso del 2018. La famiglia A spende in media 1000 euro al mese, con una varianza di 200 euro². La famiglia B spende ogni mese 8/10 della spesa della prima. La famiglia C spende ogni mese esattamente 200 euro al mese in meno della prima.

Come saranno le medie e le varianze delle spese mensili delle famiglie B e C?

- Le medie della famiglia B e C coincidono: 800 euro. La varianza delle spese mensili per la famiglia B è 128 euro² e per la famiglia C è 0
- **Le medie della famiglia B e C coincidono: 800 euro. La varianza delle spese mensili per la famiglia B è 128 euro² e per la famiglia C non varia rispetto alla famiglia A**
- Per la famiglia B, la media è 800 euro e la varianza 128 euro². La media e la varianza della famiglia C sono identiche rispetto a quelle della famiglia A
- Con i dati a disposizione non posso calcolare la media e varianza per le famiglie B e C
- Non so

Soluzione:

Nel risolvere il quesito dobbiamo dapprima fare attenzione a capire bene la differenza tra la distribuzione relativa alla famiglia B e quella relativa alla famiglia C: esse non sono identiche, infatti, come potrebbe sembrare ad un primo superficiale esame. Se per esempio la famiglia A spende nel mese x 900, la famiglia B spenderà 720 (8/10 di A) e la famiglia C 700; se la famiglia A spende nel mese y 1100, la famiglia B spenderà 880 e la famiglia C 900.

Detto questo osserviamo cosa succede alla media aritmetica quando si opera una trasformazione lineare della variabile presa in esame: $Y = aX + b$. La media diventa:

$$M_Y = aM_X + b$$

Nel caso in esame avremo per la famiglia B: $M_B = 8/10M_A = 8/10 \cdot 1000 = 800$. Per la famiglia C avremo: $M_C = M_A - 200 = 1000 - 200 = 800$. Quindi sia per B che per C la media della spesa mensile sarà di 800 euro.

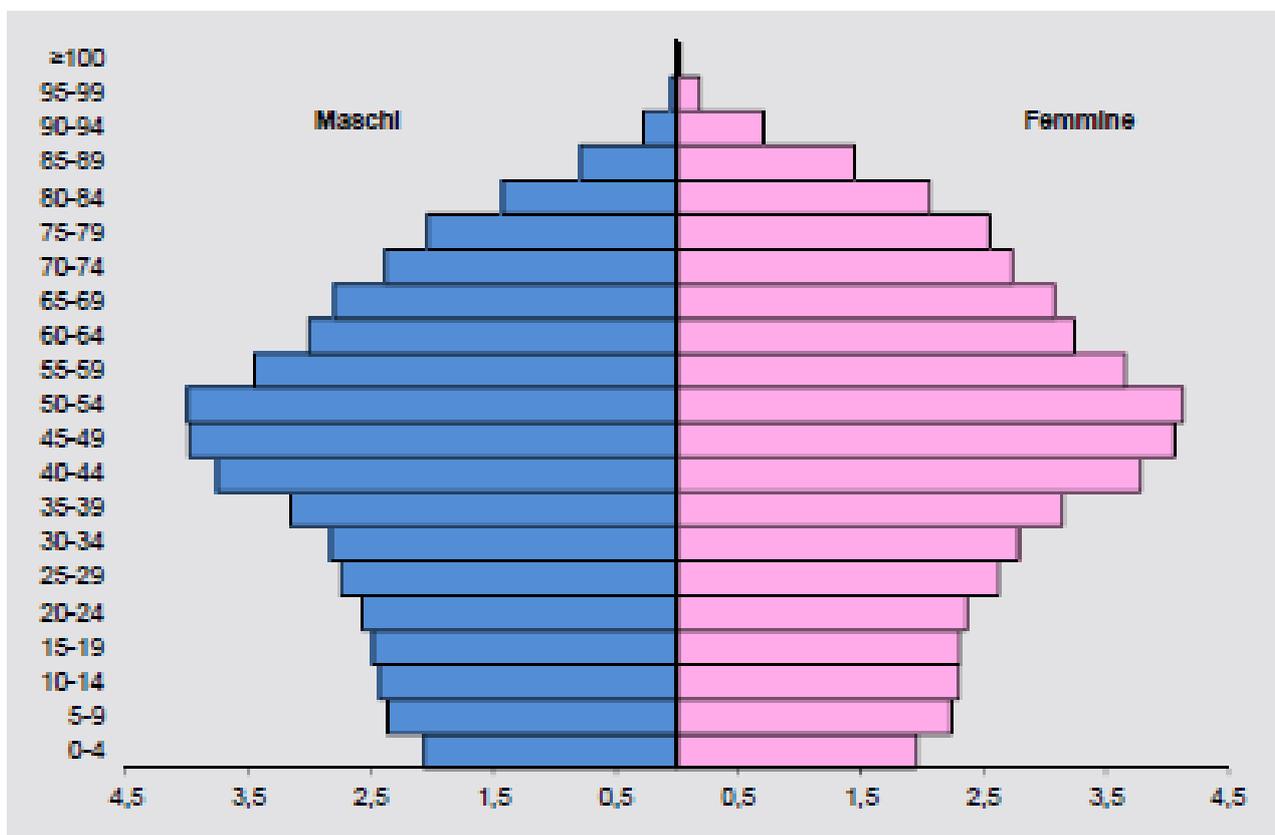
Per la varianza vale la seguente formula, sempre relativa ad una trasformazione lineare:

$$\sigma^2_{aX+b} = a^2\sigma^2_X$$

Nel nostro caso avremo, quindi, per B: $\sigma^2_B = (8/10)^2\sigma^2_A = 0,64 \cdot 200 = 128$ euro². Per la famiglia C avremo, invece: $\sigma^2_C = \sigma^2_A = 200$ euro². Per la famiglia B la varianza diventa di 128 euro², mentre per la famiglia C essa è uguale alla varianza della famiglia A. L'affermazione corretta è, quindi, la seconda. Analizzando le altre tre affermazioni, in base alle due formule citate, si può constatare come esse siano false.

QUESITO 8

Il grafico riporta la distribuzione percentuale delle età per i due sessi relative ai residenti nei comuni italiani nel 2018 (nota come “piramide delle età”; fonte ASI 2018):



Riguardo al grafico quale delle seguenti affermazioni è errata:

- Limitando l'attenzione a coloro che hanno un'età fra i 90 e i 94 anni, la probabilità di selezionare una femmina è più alta della probabilità di selezionare un maschio
- Fra tutti i residenti, la probabilità di estrarre una femmina di età compresa fra i 50 e i 54 anni è superiore alla probabilità di qualsiasi altro evento rappresentato sul grafico
- **Il grafico descrive una distribuzione percentuale, quindi non possiamo sapere se il numero di maschi con età superiore ai 60 anni è superiore o inferiore al numero di femmine di pari età**
- Dato che le righe rosa sono in genere più lunghe di quelle azzurre per le età avanzate, e sostanzialmente coincidenti o di poco inferiori per le età più basse, ci si aspetta che la media e la mediana delle età per le donne siano superiori ai corrispondenti valori per gli uomini
- Non so

Soluzione:

Le percentuali riportate nella piramide di età suddetta sono calcolate sul totale di maschi e femmine. Se osserviamo le classi di età più anziane notiamo come per ogni classe di età vi sia una preponderanza delle femmine sui maschi, dovuta ad una speranza di vita molto più alta per le donne che per gli uomini. Da tutto ciò possiamo dedurre che il numero di donne maggiori di 60 anni sia molto più alto del numero di uomini maggiori di tale età, per cui la terza affermazione è quella errata. Le altre tre affermazioni riguardanti il grafico sono, invece, vere.

QUESITO 9

Una compagnia di assicurazioni ha diviso i propri assicurati in tre gruppi: ad alto rischio incidente (15% degli assicurati), a medio rischio (50%) e a basso rischio (35%). Lo scorso anno hanno avuto un incidente il 35% degli appartenenti al primo gruppo, il 10% del secondo e il 2% del terzo. Mantenendo inalterate le cifre descritte in precedenza per l'anno corrente, quale percentuale di incidenti ci si attende fra tutti gli assicurati?

- 47%
- 50%
- 10,95%**
- 15,67%
- Non so

Soluzione:

Per rispondere al quesito dobbiamo fare una media ponderata delle persone che hanno avuto un incidente lo scorso anno (in percentuale), con i pesi dati dalla composizione percentuale dei tre gruppi di assicurati, ad alto, medio e basso rischio, ovvero:

Percentuale attesa incidenti anno corrente = $(35*15+10*50+2*35)/100 = 1095/100 = 10,95$

La percentuale corretta è, quindi, la terza.

QUESITO 10

L'insieme dei matrimoni per i quali il tribunale ha stabilito la separazione fra i coniugi nel corso del 2015, si distribuisce come nella tabella seguente rispetto al titolo di studio della moglie e del marito (fonte dati.istat.it):

Titolo di studio del marito	Titolo di studio della moglie				Totale
	Licenza di scuola elementare, nessun titolo di studio	Licenza di scuola media	Diploma	Laurea e post-laurea	
Licenza di scuola elementare, nessun titolo di studio	1717	1544	432	44	3737
Licenza di scuola media	1196	16793	8950	1340	28279
Diploma	299	5167	21100	4917	31483
Laurea e post-laurea	19	468	3461	6591	10539
Totale	3231	23972	33943	12892	74038

Se si pesca a caso una coppia di coniugi separati, si determini la probabilità dei seguenti eventi:

A=nella coppia selezionata, il marito possiede almeno la licenza di scuola media

B=fra le coppie dove il marito possiede la licenza di scuola media, la moglie ha un titolo superiore o uguale al diploma

- $P(A)=94,95\%$, $P(B)=51,18\%$
- $P(A)=94,95\%$, $P(B)=36,39\%$**
- $P(A)=38,19\%$, $P(B)=67,30\%$
- $P(A)=32,38\%$, $P(B)=22,68\%$
- Non so

Soluzione:

Applicando la definizione classica di probabilità, ovvero quella che la vede come rapporto tra il numero dei casi favorevoli e il numero dei casi possibili, abbiamo per l'evento A:

$$P(A) = (28279+31483+10539)/74038*100 = 70301/74038*100 = 94,95$$

Per determinare la probabilità dell'evento A dobbiamo sommare tutti le coppie di coniugi separati in cui i mariti possiedono la licenza di scuola media, il diploma o la laurea e dividere tale somma per il totale delle coppie, moltiplicando per cento il risultato ottenuto. Per la probabilità dell'evento B abbiamo invece:

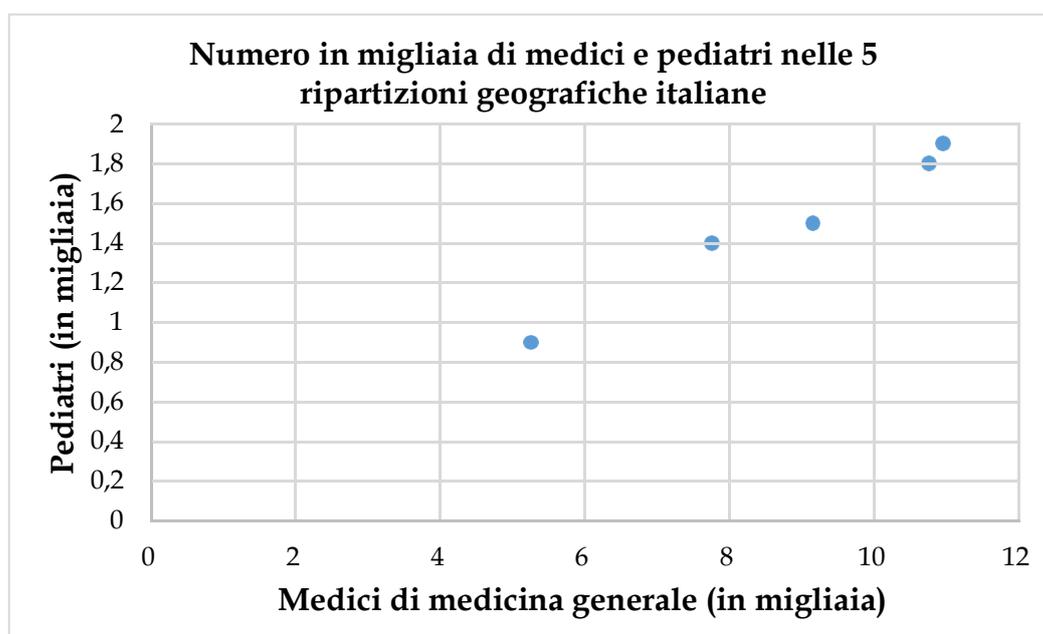
$$P(B) = (8950+1340)/28279*100 = 10290/28279*100 = 36,39$$

Nel caso dell'evento B dobbiamo far riferimento alla sola riga relativa alle coppie di coniugi separati in cui il marito ha la licenza di scuola media. In tale riga dobbiamo sommare il dato relativo alle coppie in cui la moglie ha il diploma o la laurea e dividere tale somma per il totale delle coppie in cui il marito ha la licenza di scuola media, moltiplicando il risultato per cento. Grazie ai risultati ottenuti possiamo affermare che la seconda coppia di probabilità è quella corretta.

QUESITO 11

Nel 2016 i medici di medicina generale e i pediatri di libera scelta disponibili in Italia (dati in migliaia) si distribuivano nelle 5 ripartizioni geografiche italiane (fonte dati.istat.it) come dalla tabella seguente. Si fornisce anche una rappresentazione grafica:

Territorio	Medici di medicina generale	Pediatri di libera scelta
Nord-ovest	10,8	1,8
Nord-est	7,8	1,4
Centro	9,2	1,5
Sud	11	1,9
Isole	5,3	0,9



Si determini il valore del coefficiente di correlazione fra il numero di medici di medicina generale e i pediatri di libera scelta:

- è quasi nullo e positivo, indicando una leggera associazione lineare concorde fra i due caratteri
- è quasi pari a 1 indicando una fortissima associazione lineare concorde fra i due caratteri**
- è quasi nullo ma negativo, indicando una leggera associazione lineare discorde
- è quasi pari a -1 indicando una fortissima associazione lineare discorde fra i due caratteri
- Non so

Soluzione:

Il coefficiente di correlazione lineare di Pearson ρ_{XY} tra le variabili X e Y (rispettivamente il numero di medici di medicina generale ed il numero di pediatri di libera scelta) è dato da:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Nella suddetta formula σ_{XY} indica la covarianza delle due variabili e σ_X e σ_Y rispettivamente la deviazione standard di X e di Y. Per calcolare la covarianza tra le due variabili dobbiamo utilizzare la formula già utilizzata per l'esercizio 1:

$$\sigma_{X,Y} = \frac{1}{N} \sum (x_i - \bar{x})(y_i - \bar{y})$$

Per calcolare la deviazione standard di X possiamo utilizzare la seguente formula:

$$\sigma_X = \sqrt{\frac{\sum_i (x_i - \bar{x})^2}{N}}$$

Si avrà poi l'analogia formula per la deviazione standard di Y:

$$\sigma_Y = \sqrt{\frac{\sum_i (y_i - \bar{y})^2}{N}}$$

Dovremo quindi, prima di calcolare covarianza e deviazioni standard, calcolare le medie di X e di Y.

Dalle suddette formule si può facilmente intuire la complessità dei calcoli per ottenere il coefficiente di correlazione lineare. Possiamo, però, ovviare a tutti i calcoli citati osservando attentamente la tabella e soprattutto il grafico forniti nell'esercizio. Osserviamo, infatti, come i due caratteri siano fortemente correlati positivamente tra loro, com'è intuitivo che sia: dal grafico si intuisce che una eventuale retta di regressione passerebbe molto vicino ai punti del grafico e che una retta che congiungesse tali punti si sovrapporrebbe quasi alla suddetta retta di regressione. Possiamo dedurre da tutto ciò che la seconda risposta è quella corretta e che vi sia tra le due variabili una fortissima associazione lineare concorde e che quindi il coefficiente di correlazione lineare abbia un valore vicino ad uno. Tutte le altre risposte sono false, com'è facile stabilire osservando attentamente la tabella e soprattutto il grafico.

QUESITO 12

L'indagine Istat multiscopo sugli aspetti della vita quotidiana condotta nel 2017 riporta, per i collettivi delineati dalle diverse condizioni professionali, i valori percentuali relativi a coloro che si dichiarano in buona salute e a coloro che dichiarano di soffrire di una malattia cronica:

Condizione nella professione	% di persone di 15 anni e più in buona salute	% di persone di 15 anni e più con almeno una malattia cronica
Occupato	78,4	31,9
Casalinga/o	51,7	58,1
Studente	91,2	17,9
Ritirato-a dal lavoro	36,4	79,5
In altra condizione	43,6	60,4
Disoccupato alla ricerca di nuova occupazione	68	36,3
In cerca di prima occupazione	83,8	19,1

Si noti che si può essere in buona salute anche in presenza di una malattia cronica (ad esempio in buona salute ma affetto da ipertensione). Si selezioni l'affermazione corretta:

- Le percentuali nella tavola sono sbagliate perché la somma dei dati per riga non fa 100
- Fra le casalinghe, la percentuale di coloro che non sono in buona salute è il 48,3%**
- Le persone di 15 anni e più con almeno una malattia cronica a prescindere dalla condizione professionale sono il 43,3%
- Gli studenti di 15 anni e più sono il 54,6%
- Non so

Soluzione:

Se calcoliamo il complemento a 100 per la percentuale di casalinghe in buona salute otteniamo:

Percentuale di casalinghe che **non** sono in buona salute = $100 - 51,7 = 48,3$

Quindi l'affermazione corretta è la seconda. Tutte le altre affermazioni sono false, com'è facile verificare osservando i dati della tabella. In particolare, è falsa la prima affermazione: viene infatti specificato nel testo dell'esercizio che si può essere in buona salute anche in presenza di una malattia cronica, per cui la somma delle due percentuali per riga non deve essere 100. Ci sarà, infatti, una percentuale, variabile a seconda dei casi, data dalla differenza tra la somma delle due percentuali per riga e 100, di persone che ogni volta si dichiarano in buona salute pur in presenza di una malattia cronica.

QUESITO 13

La tabella presenta i dati dell'indagine sui redditi (EU-SILC) nel 2016 per le famiglie italiane, relativi ai redditi medi e alla percentuale di famiglie in cui risiede un minore, per fonte principale del reddito familiare (fonte dati.istat.it):

Fonte principale di reddito familiare	Reddito medio	% di famiglie in cui risiede un minore
Lavoro dipendente	36088	71,9
Lavoro autonomo	36716	17,9
Pensioni e trasferimenti pubblici	30263	7,8
Altro	17069	2,4

Quale di queste affermazioni è vera:

- La mediana della fonte principale di reddito è "lavoro dipendente"
- Il reddito medio complessivo di tutte le famiglie con un minore è di 30034 euro
- Il reddito medio complessivo di tutte le famiglie con un minore è di 35289,61 euro**
- Non è possibile calcolare la moda della fonte principale di reddito familiare per le famiglie con un minore perché in base alla distribuzione percentuale dovrebbe essere "lavoro dipendente" ma rispetto al reddito medio dovrebbe essere "lavoro autonomo"
- Non so

Soluzione:

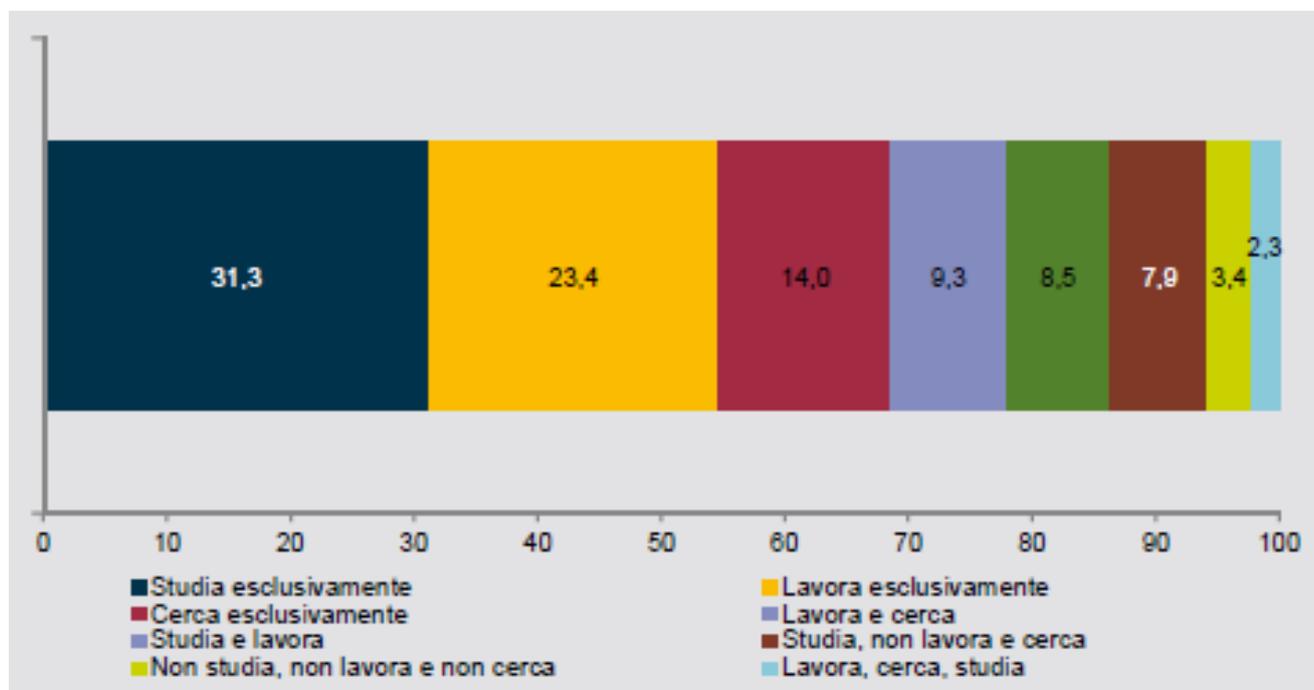
Per calcolare il reddito medio complessivo di tutte le famiglie con un minore dobbiamo calcolare una media ponderata dei redditi delle 4 tipologie di famiglie, avente come pesi le percentuali di famiglie con un minore; per ottenere tale media ponderata dobbiamo sommare i prodotti di ciascun reddito medio per la percentuale relativa di famiglie con un minore e dividere la somma complessiva per 100. Ovvero:

Reddito medio complessivo di tutte le famiglie con un minore = $[(36088 \cdot 71,9) + (36716 \cdot 17,9) + (30263 \cdot 7,8) + (17069 \cdot 2,4)] / 100 = 35289,61$

Quindi la risposta corretta è la terza. Analizzando i dati in tabella si può evincere come le altre tre affermazioni dell'esercizio siano tutte false.

QUESITO 14

Il seguente grafico, tratto dall'Annuario Statistico Italiano 2018, riporta la condizione occupazionale nel 2015 di coloro che si sono diplomati nel 2011:



Cosa si può dire di questa distribuzione:

- Ordinando e cumulando le percentuali indicate nella figura, la mediana è 23,4%
- **Il carattere è qualitativo non ordinabile, la moda è "Studia esclusivamente"**
- Il carattere è quantitativo e posso calcolare la media, che risulta 12,50%
- I valori rappresentati sul grafico non corrispondono a una distribuzione di frequenze percentuali
- Non so

Soluzione:

Il carattere "condizione occupazionale nel 2015 di coloro che si sono diplomati nel 2011" è certamente qualitativo non ordinabile (non è possibile stabilire un ordine di priorità tra le varie modalità) e la modalità che presenta la percentuale più alta, 31,3%, è "studia esclusivamente"; essa è quindi anche la moda della distribuzione. La risposta corretta è quindi la seconda. Osservando il grafico si può facilmente stabilire che le altre tre risposte fornite sono tutte false.

QUESITO 15

Gianna sa che se l'autobus delle 8 non arriva in ritardo sarà in grado di entrare a scuola in tempo. Sa anche che quando il tempo è bello, l'autobus fa ritardo solo nel 5% dei casi mentre con la pioggia l'autobus ritarda nel 50% dei casi. Le previsioni del tempo hanno dichiarato che domani pioverà con una probabilità del 60%. Con quale probabilità Gianna riuscirà ad entrare a scuola in tempo?

- 32%
- 50%
- 68%**
- 30%
- Non so

Soluzione:

Per ottenere la probabilità richiesta dobbiamo effettuare i seguenti semplici calcoli:

Probabilità che Gianna entri a scuola in tempo = $[(95 \cdot 40) + (50 \cdot 60)] / 100 = 6800 / 100 = 68$

Ovvero dobbiamo sommare il prodotto della probabilità che l'autobus arrivi puntuale col tempo bello (95%) per la probabilità che sia bel tempo (40%) al prodotto della probabilità che l'autobus arrivi puntuale con la pioggia (50%) per la probabilità che piova (60%); dobbiamo poi dividere il tutto per 100, ottenendo così la probabilità richiesta. La risposta corretta è dunque la terza.

QUESITO 16

Un incendio ha distrutto un archivio pieno di documenti. Uno di questi contiene una tabella di dati che riporta i dipendenti dell'ufficio per titolo di studio. Purtroppo, ora si legge solo parzialmente:

Titolo di studio	Numero di addetti	Frequenza percentuale
Senza titolo e licenza elementare	5	20
Licenza di scuola media inferiore	?	?
Diploma di scuola media superiore	10	?
Laurea, diploma universitario, corsi post-laurea	5	?
Totale	?	?

Qual è il valore mediano del titolo di studio?

- Non è possibile calcolarlo, mancano i dati
- Licenza di scuola media inferiore
- Diploma di scuola media superiore**
- Laurea, diploma universitario, corsi post-laurea
- Non so

Soluzione:

Per risolvere l'esercizio dobbiamo dapprima calcolare il totale dei dipendenti, mancante dalla tabella. Per farlo possiamo far ricorso ad una semplice proporzione:

$$5 : 20 = x : 100$$

Ovvero la frequenza assoluta 5, riferita ai dipendenti senza titolo o con licenza elementare, sta a 20, la corrispondente frequenza percentuale, come x, il totale dei dipendenti, sta a 100. Da essa si ricava facilmente:

$$x = 5 \cdot 100 / 20 = 25$$

Una volta determinato il totale possiamo calcolare agevolmente il numero di dipendenti con licenza di scuola media inferiore:

$$\text{Dipendenti con licenza di scuola media inferiore} = 25 - (5 + 10 + 5) = 5$$

Abbiamo ora la distribuzione completa delle frequenze assolute. Poiché la metà di 25 è 12,5 e cumulando tali frequenze assolute, possiamo dedurre che il valore mediano corrisponde alla modalità "Diploma di scuola media superiore" (prima di tale modalità cumulando le frequenze abbiamo 10 dipendenti, dopo di essa cumulando abbiamo 20 dipendenti ed il valore centrale 12,5 è proprio compreso tra tali due valori).

QUESITO 17

In una squadra di pallavolo femminile di 6 giocatrici le altezze sono:

1,72m, 1,80m, 1,82m, 1,82m, 1,84m, 1,86m

Le giocatrici più bassa e più alta del sestetto vengono sostituite da due altre giocatrici che hanno altezze diverse dalle loro, ma che fanno rimanere invariata la mediana e la media aritmetica della distribuzione. Quali dei seguenti valori delle altezze delle sostitute sono compatibili con quest'ultima condizione?

- Il problema non ha soluzione
- 1,78m, 1,92m
- 1,69m, 1,89m**
- Qualsiasi valore fra 1,73 e 1,82 per la più bassa e fra 1,82 e 1,86 per la più alta
- Non so

Soluzione:

Prendiamo in considerazione la terza possibile risposta al quesito posto, quella relativa alle altezze di 1,69 m e di 1,89 m. Essa lascia invariata la mediana (che è anche in tal caso uguale a 1,82 m, semisomma dei due valori centrali, tra loro identici) ma lascia invariata anche la media aritmetica, visto che la somma complessiva delle altezze non cambia nei due casi: 1,69 m è più basso di 3 cm rispetto a 1,72 m e 1,89 m è più alto di 3 cm rispetto a 1,86 m. Possiamo verificarlo attraverso semplici calcoli:

Media iniziale = $(1,72+1,80+1,82+1,82+1,84+1,86)/6 = 10,86/6 = 1,81$

Media finale = $(1,69+1,80+1,82+1,82+1,84+1,89)/6 = 10,86/6 = 1,81$

Tale risposta è quindi quella corretta. Le altre tre risposte sono tutte false. In particolare, i valori relativi alla seconda e alla quarta risposta lasciano invariata la mediana della distribuzione ma non la sua media aritmetica.

QUESITO 18

Tre scatole A, B e C contengono lampade prodotte da una certa fabbrica di cui alcune difettose. A contiene 1000 lampade con il 10% di esse difettose, B ne contiene 1500 con il 15% difettose e C ne contiene 500 con il 5% difettose. Si sceglie una scatola a caso e si estrae a caso una lampada.

Quale è la probabilità che essa sia difettosa?

- 0,05
- 0,10**
- 0,15
- 0,30
- Non so

Soluzione:

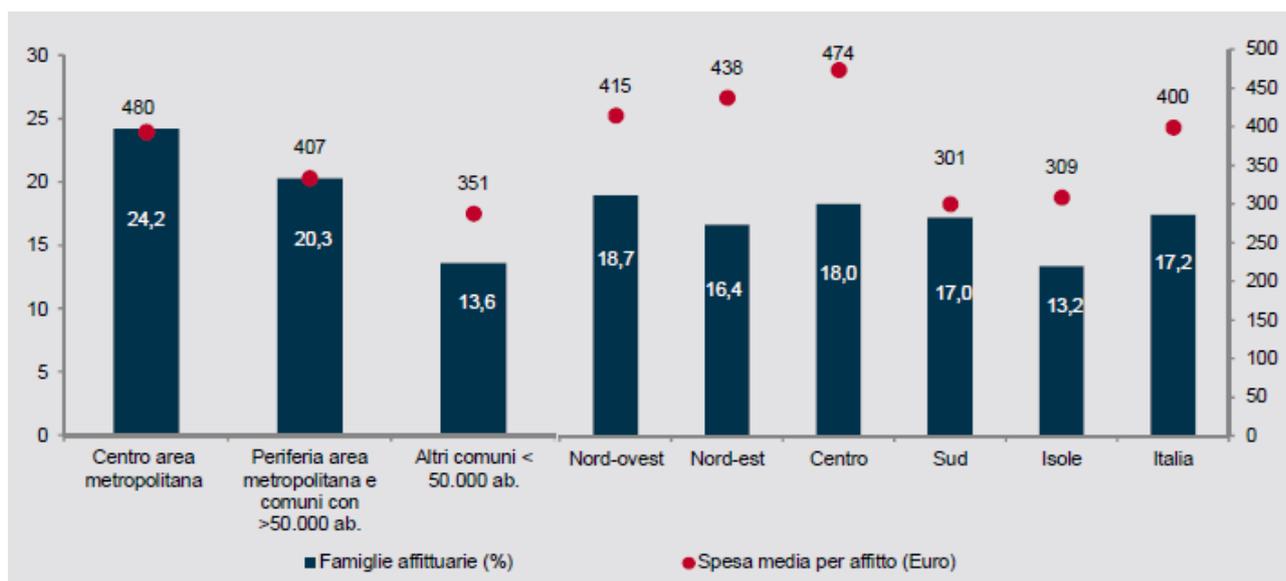
Per determinare la probabilità richiesta dobbiamo effettuare i seguenti semplici calcoli:

Probabilità che la lampada sia difettosa = $1/3 \cdot (0,10 + 0,15 + 0,05) = 1/3 \cdot 0,30 = 0,10$

Ovvero si moltiplica la probabilità di estrarre una scatola (1/3) per la somma delle tre diverse probabilità di estrarre una lampada difettosa. L'operazione effettuata equivale a calcolare una media semplice delle tre probabilità, senza far riferimento ai pesi del numero delle lampade per ogni scatola. Tale modo di operare è dovuto al fatto che effettuando l'estrazione della lampada in due stadi (prima la scelta casuale della scatola e poi l'estrazione casuale della lampada) si "neutralizza" la diversa numerosità delle lampade per ogni scatola, ai fini del calcolo della probabilità richiesta. La risposta corretta è, quindi, la seconda.

QUESITO 19

L'Indagine sulle spese delle famiglie italiane fornisce la percentuale delle famiglie in affitto e la spesa media per gli affitti (in euro) delle famiglie. Il grafico presenta i dati relativi alle famiglie suddivisi per "dimensione demografica del comune di residenza" e "ripartizione geografica di residenza" (fonte: Annuario Statistico Italiano, 2018):



Quali delle seguenti affermazioni è vera:

- Ha senso costruire una retta di regressione con i valori rappresentati dai punti rossi nel grafico
- Il valore mediano delle spese medie in affitto è 407 euro
- La media aritmetica semplice delle percentuali di famiglie affittuarie nelle 5 ripartizioni geografiche equivale alla percentuale di famiglie affittuarie in tutta Italia
- Allontanandosi dal centro delle aree metropolitane, chi vuole andare ad abitare in affitto si attende costi via via più bassi, mentre è più facile avere come vicine famiglie proprietarie di casa**
- Non so

Soluzione:

Osservando il grafico, ed in particolare le sue prime tre colonne, relative alla dimensione demografica del comune di residenza, si nota come, allontanandosi dal centro dell'area metropolitana verso i comuni più piccoli, la percentuale di famiglie affittuarie diminuisca e con essa anche la spesa media per affitto, com'è naturale che sia. La quarta affermazione è, quindi, quella corretta. Sempre osservando il grafico si può stabilire che le altre tre affermazioni sono tutte false.

QUESITO 20

La seguente tabella rappresenta il numero (in migliaia) di fumatori di 14 anni e più per titolo di studio e numero di sigarette fumate al giorno espresso in classi per l'anno 2017 (fonte: dati.istat.it):

Titolo di studio	Persone di 14 anni e più fumatori per sigarette fumate - Anno 2017, in migliaia			
	fino a 5 sigarette	da 6 a 10 sigarette	da 11 a 20 sigarette	oltre 20 sigarette
Licenza di scuola elementare, nessun titolo di studio	15	16	61	6
Licenza di scuola media	257	466	595	25
Diploma	502	709	670	56
Laurea e post-laurea	263	187	137	5

Si assuma che l'ultima classe sia chiusa al valore "40 sigarette"

Quale delle seguenti affermazioni è vera:

- **Il numero medio di sigarette fumate da chi ha la licenza di scuola elementare o nessun titolo è 13,3**
- Al crescere del titolo di studio cresce il numero medio di sigarette fumate
- Il numero medio di sigarette fumate da chi ha la laurea o un diploma post-laurea è 10,3
- Non è possibile calcolare la media di un carattere espresso in classi
- Non so

Soluzione:

Per rispondere alla domanda posta nell'esercizio dobbiamo dapprima calcolare i quattro valori centrali delle classi, che sono rispettivamente: 3, 8, 15,5 e 30. Possiamo ora calcolare il primo numero medio di sigarette fumate, quello relativo ai fumatori con licenza di scuola elementare o nessun titolo di studio, attraverso i seguenti semplici calcoli:

Numero medio di sigarette fumate dai fumatori con licenza elementare o nessun titolo = $(3 \cdot 15 + 8 \cdot 16 + 15,5 \cdot 61 + 30 \cdot 6) / 98 = 1298,5 / 98 = 13,3$

Nei calcoli suddetti 98 (migliaia) è dato dalla somma di tutti i fumatori con licenza elementare o nessun titolo di studio. Analogamente si può calcolare il numero medio di sigarette per le altre tre tipologie di fumatori, ottenendo per i fumatori con licenza di scuola media un numero medio di 10,8 sigarette, per quelli con diploma di 9,9 sigarette e per quelli con laurea o diploma post-laurea di 7,7 sigarette. L'affermazione corretta è quindi la prima, le altre sono tutte false; in particolare rispetto alla seconda, e tenendo conto dei numeri medi di sigarette fumate dalle 4 tipologie di fumatori, si può dire esattamente il contrario: al crescere del titolo di studio **decresce** il numero medio di sigarette fumate, come si può riscontrare semplicemente osservando la tabella, anche senza effettuare i calcoli suddetti.