



XV edizione - Anno 2025

Prove individuali

13-16 gennaio 2025

Classi III e IV

Quesiti

A cura di: Francesco Michele Mortati

Quesiti di: Flavio Foschi, Rita Lima, Clara Romano, Mauro Scanu

Progetto grafico: Bruna Tabanella

olimpiadi-statistica@istat.it

Secondo la pubblicazione Eurostat sulla digitalizzazione Ue, nel 2022 l'e-commerce è stato praticato nelle seguenti proporzioni, tenendo conto della dimensione delle imprese (la distinzione tra "Grandi Imprese (G.I.)" e "Piccole e Medie Imprese (P.M.I.)" riguarda la dimensione aziendale in termini di numero di addetti):

G.I.	P.M.I.	TOTALE IMPRESE
0,46	0,22	0,23

Quale affermazione tra le seguenti *non* è supportata dai dati?

- **Il numero assoluto di P.M.I. che hanno svolto e-commerce è inferiore al corrispondente numero assoluto di G.I.**
- Le G.I. hanno una propensione all'e-commerce maggiore delle P.M.I.
- La proporzione delle P.M.I. che svolgono e-commerce è circa la metà della proporzione relativa alle G.I.
- La proporzione delle imprese (di qualsiasi dimensione) che praticano e-commerce è simile alla proporzione delle P.M.I. che praticano l'e-commerce
- Non so

Soluzione

È errata la risposta n. 1: le informazioni fornite "condizionatamente al numero di addetti" non riguardano quantità espresse nei livelli ma rapporti tra il numero di imprese aventi una certa dimensione che praticano e-commerce e il numero totale di imprese di quella dimensione; quindi, non si possono riferire a numeri assoluti di imprese.

Nella stagione 2023/2024, la copertura del vaccino antinfluenzale della popolazione italiana si è assestata intorno al 20%. Ipotizzando che il vaccino abbia un'efficacia del 90% (ovvero che protegga dall'influenza, in media, 90 vaccinati su 100), qual è la probabilità che un italiano scelto a caso risulti essere vaccinato ed ammalato?

- 0,10
- 0,12
- **0,02**
- 0,30
- Non so

Soluzione

La risposta corretta è la n. 3.

Sia A l'evento "l'individuo è stato vaccinato" e B l'evento "l'individuo si è ammalato". Sappiamo che $P(A) = 0,2$ e $P(B^c | A) = 0,9$.

Avremo: $P(B | A) = 1 - P(B^c | A) = 0,1$ (probabilità di ammalarsi essendosi vaccinati)

$P(A \cap B) = P(B | A) * P(A) = 0,1 * 0,2 = 0,02$ (ovvero il 2%)

Ovvero la probabilità che l'italiano scelto a caso sia vaccinato ed ammalato è del 2%.

La variabile X assume i quattro valori seguenti:

X	1	2	3	6
-----	---	---	---	---

La varianza della variabile X è pari a:

- 1,9
- 3,5
- 5
- 3
- Non so

Soluzione

X	1	2	3	6	Media =	3
$(X - \text{media})^2$	4	1	0	9	Varianza =	3,5

La risposta corretta è la n. 2: infatti la somma dei quadrati degli scarti dalla media vale 14. Quindi la varianza è uguale a $14/4 = 3,5$.

Secondo la pubblicazione Eurostat sulla digitalizzazione nella Ue, nel 2023 i social media maggiormente utilizzati dalle imprese sono stati:

Social media	Facebook e LinkedIn	YouTube e Instagram	Corporate Blogs
Frequenza % dei casi	59	31	10

La mediana della suddetta distribuzione è:

- 59
- “Facebook e LinkedIn”
- 31
- **Non calcolabile**
- Non so

Soluzione

È corretta la risposta n. 4: il calcolo della mediana richiede che le modalità del carattere siano almeno ordinabili, mentre la variabile in esame è qualitativa sconnessa e quindi non ordinabile. In tal caso l'unico indice di posizione calcolabile è la moda.

L'Annuario Statistico Italiano, edizione 2023, a p. 756, riporta la distribuzione per classe di addetti delle imprese che hanno svolto attività di ricerca e sviluppo nel triennio 2018-2020. Considerando la tabella:

Classe di addetti	10-49	50-249
Numero imprese	60.742	9.799

Il numero medio di addetti nelle imprese, con dimensione compresa tra 10 e 249 e che hanno svolto attività di ricerca e sviluppo nel triennio 2018-2020, è:

- 46,17
- 26,79
- 72,30
- 58,88
- Non so

Soluzione

È corretta la risposta n. 1: in assenza di altre informazioni il valore centrale di ciascuna classe viene assunto come valore rappresentativo, al fine di procedere con il calcolo della media, ponderata per mezzo delle corrispondenti frequenze.

Ovvero, passando ai calcoli, si determinano dapprima i valori centrali delle due classi di età, come semisomme degli estremi delle due classi di addetti:

$$S(10-49) = (10+49)/2 = 29,5$$

$$S(50-249) = (50+249)/2 = 149,5$$

Si moltiplicano successivamente tali valori centrali per le rispettive frequenze:

$$29,5 * 60.742 = 1.791.889$$

$$149,5 * 9.799 = 1.464.951$$

Si sommano, infine, tali prodotti e si divide la somma per il numero totale di imprese:

$$M = (1.791.889 + 1.464.951) / (60.742 + 9.799) = 3.256.840 / 70.541 = 46,17$$

I dati Eurostat sulla digitalizzazione nei Paesi Ue riportano, per il 2023, la seguente situazione, rispetto al livello base di digitalizzazione delle imprese (la distinzione tra “Grandi Imprese (G.I.)” e “Piccole e Medie Imprese (P.M.I.)” riguarda la dimensione aziendale in termini di numero di addetti):

Variabile X	Variabile Y		
	Piccole e Medie Imprese	Grandi Imprese	Tutte le Imprese
Livello base non conseguito	0,42	0,09	0,41
Livello base conseguito	0,58	0,91	0,59

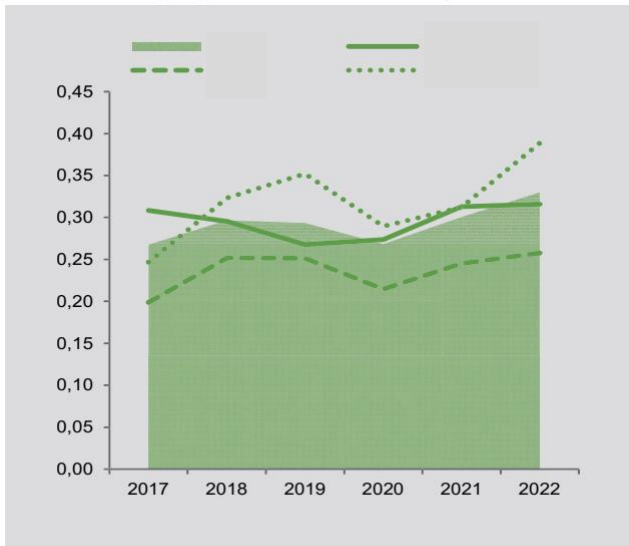
Questa tabella mostra:

- Le frequenze marginali di X (livello base di digitalizzazione)
- **La distribuzione di frequenza di X condizionata da Y**
- Le frequenze marginali di Y (dimensione delle imprese)
- La distribuzione di frequenza di Y condizionata da X
- Non so

Soluzione

È corretta la risposta n. 2: la somma per colonna delle frequenze relative è pari a 1, mentre la somma lungo le righe assume valori diversi da 1, quindi la tabella riporta la distribuzione di frequenza di X condizionata da Y.

Il seguente diagramma, ricavato dalla figura 15.5b, p. 166 del rapporto SDGs 2024 dell'Istat, oscurandone le etichette “Nord”, “Centro”, “Mezzogiorno” e “Italia”, mostra l'evoluzione nel tempo della quota di superficie territoriale impermeabilizzata da copertura artificiale:



Fonte: ISPRA

Quale “curva”, tra le quattro rappresentate, riguarda l'Italia?

- La linea continua
- La linea tratteggiata con punti
- **Il profilo dell'area colorata in verde**
- La linea tratteggiata con segmenti
- Non so

Soluzione

È corretta la risposta n. 3: il valore riferito all'Italia è necessariamente una media di quelli che caratterizzano le aree geografiche “Nord”, “Centro” e “Mezzogiorno”; dunque, rispetto a questi ultimi deve verificare la c.d. proprietà di “internalità”; dal grafico si vede chiaramente che il profilo dell'area verde non eccede mai i valori più alti e più bassi dei diversi anni, diversamente da quanto accade per le tre linee.

In riferimento alla distribuzione di una variabile quantitativa, con il termine “deviazione standard” si indica:

- La somma dei valori assunti dalla variabile
- La somma dei quadrati degli scarti dalla media aritmetica
- **La radice quadrata del rapporto tra la somma degli scarti, elevati al quadrato, dalla media ed il numero di unità statistiche**
- Il rapporto tra la somma degli scarti dalla media aritmetica ed il numero di unità statistiche
- Non so

Soluzione

È corretta la risposta n. 3: la deviazione standard, o scostamento quadratico medio, è data dalla radice quadrata del rapporto tra la sommatoria degli scarti dalla media aritmetica, ogni volta elevati al quadrato, ed il numero di unità statistiche.

Supponiamo che in una classe la statura media di tutti gli alunni e le alunne sia 168 cm. Le alunne sono alte in media 165 cm. Qual è la statura media in cm degli alunni, supponendo che nella classe ci siano 15 alunni e 10 alunne?

- 173
- **170**
- 176
- 167
- Non so

Soluzione

Per la proprietà associativa, la statura media di tutti gli alunni e alunne è la media ponderata delle stature medie dei due gruppi, con pesi di ponderazione pari alle numerosità di tali gruppi. Indicando con x la statura media degli alunni, a noi sconosciuta, vale la relazione $(x*15+165*10)/25 = 168$, da cui, risolvendo per x , si ottiene $x = 170$.

x = statura media degli alunni

$$(x*15+165*10)/25 = 168$$

$$x*15+165*10 = 4.200$$

$$x*15+1.650 = 4.200$$

$$x*15 = 4.200 - 1.650 = 2.550$$

$$x = 2.550/15 = 170$$

Come verifica del risultato ottenuto si può ricalcolare la media complessiva, partendo dalle medie dei due gruppi:

$x(i)$	$n(i)$	$x(i)*n(i)$
165	10	1.650
170	15	2.550
Totale	25	4.200
	Media =	168

L'Annuario Statistico Italiano, edizione 2023, p. 751, riporta i dati relativi alle spese in Ricerca e Sviluppo sostenute da alcuni settori nel 2021, qui arrotondate al milione di euro:

Istituzioni pubbliche	Imprese	Università
3.629	15.645	6.232

Sapendo che la media per settore delle spese in Ricerca e Sviluppo è 8.502 (milioni di euro), la deviazione standard arrotondata (in milioni di euro) è pari a:

- **5.161**
- 23.746.129
- 51.022.449
- 5.152.900
- Non so

Soluzione

È corretta la risposta n. 1: si noti, infatti, che il valore della deviazione standard è l'unico avente un ordine di grandezza comparabile con quello delle intensità osservate.

Passando ai calcoli per determinare la deviazione standard (o scostamento quadratico medio), si calcolano dapprima i quadrati dei 3 scostamenti dalla media:

$$S_1^2 = (3.629 - 8.502)^2 = 23.746.129$$

$$S_2^2 = (15.645 - 8.502)^2 = 51.022.449$$

$$S_3^2 = (6.232 - 8.502)^2 = 5.152.900$$

La deviazione standard è data dalla radice quadrata del rapporto tra la somma dei tre scostamenti al quadrato ed il numero di tali scostamenti:

$$Ds = \sqrt{\frac{1}{3}(23.746.129 + 51.022.449 + 5.152.900)} = \sqrt{26.640.493} = 5.161$$

L'Annuario Statistico Italiano, edizione 2023, p. 756, mostra la distribuzione delle imprese, nel settore delle costruzioni, che svolgono attività innovativa, distinte per classe di addetti (anni 2018-2020). Si consideri la seguente tabella riassuntiva:

Addetti	10-49	50-249	250 e oltre
Numero imprese	19.938	1.281	98

La classe mediana del numero di addetti è pari a:

- 10-49
- 50-249
- 250 e oltre
- Non è calcolabile
- Non so

Soluzione

È corretta la risposta n. 1. La classe mediana è individuata dall'intervallo 10-49, dato che tale intervallo comprende il primo 93,5% delle osservazioni. Per calcolare tale percentuale dobbiamo rapportare il numero di imprese della classe 10-49 addetti (19.938) al totale delle imprese (21.317) e moltiplicare tale rapporto per 100.

L'Annuario Statistico Italiano, edizione 2023, p. 364, riporta le statistiche sugli occupati temporanei condizionatamente all'area geografica, nel 2021. Dai dati si ottiene la seguente tabella di frequenze relative, che riporta gli occupati maschi sul totale di maschi e femmine, nella quale è stata oscurata una cella:

Occupati maschi	Nord e Centro	Sud e Isole
	0,62	****

Quale frequenza relativa ci si dovrebbe attendere in corrispondenza della colonna “Sud e Isole”, se i caratteri “Sesso degli occupati” e “Collocazione geografica” fossero indipendenti?

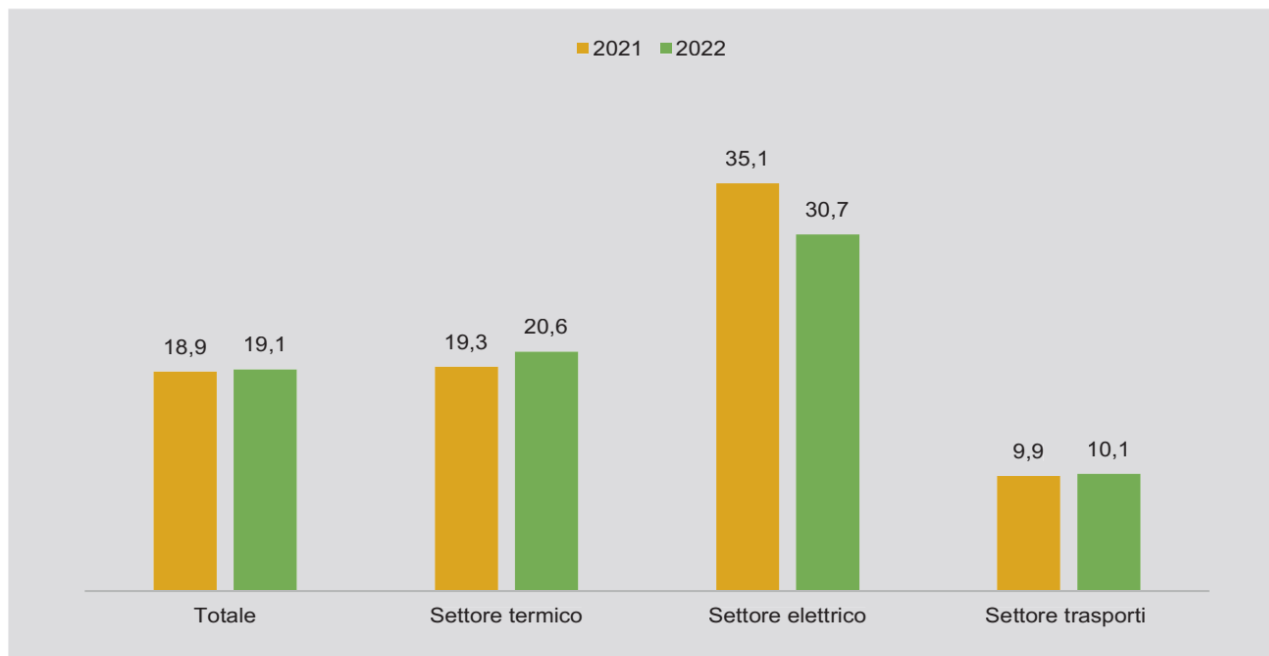
- In assenza di ulteriori informazioni non è possibile rispondere al quesito
- Un valore prossimo a 0,99
- **Un valore pari a 0,62**
- Un valore prossimo a 0,01
- Non so

Soluzione

È corretta la risposta n. 3: se il condizionamento esercitato dalla collocazione geografica rispetto al sesso degli occupati fosse assente, sarebbe lecito aspettarsi frequenze relative degli occupati maschi sul totale approssimativamente identiche nelle diverse aree geografiche.

Il rapporto SDGs 2024 dell'Istat, p. 88, mostra il seguente grafico:

Figura 7.3 - Quota di energia da fonti rinnovabili sui consumi complessivi di energia, per settore. Anni 2021-2022
(valori percentuali)



Fonte: GSE S.p.A. - Gestore dei Servizi Energetici; Terna S.p.A.

Rispetto alla variabile “Quota di energia da fonti rinnovabili” nel 2022, classificata nei settori “termico”, “elettrico” e “trasporti”, quale indice di posizione è possibile calcolare?

- Nessuno
- Mediana
- Media
- **Moda**
- Non so

Soluzione

È corretta la risposta n. 4: il carattere è di tipo qualitativo sconnesso; dunque, l'unico indice di posizione calcolabile è la moda.

Per analizzare in che misura la statura dei figli dipende da quella dei padri ha senso calcolare:

- **Il coefficiente di regressione lineare**
- Le mode
- Il coefficiente di correlazione lineare
- Le varianze
- Non so

Soluzione

È corretta la risposta n. 1, perché si tratta di una relazione di dipendenza, considerato che la statura dei padri è uno dei fattori che influenzano la statura dei figli. Non ha senso calcolare il coefficiente di correlazione, perché quest'ultimo si usa per fenomeni interdipendenti.

Secondo il rapporto SDGs 2024 dell'Istat, p. 61, “Nel 2023, aveva competenze digitali di base il 45,9% delle persone di 16-74 anni [...], con una quota più elevata tra gli uomini (47,4%), tra i giovani tra 20 e 24 anni (61,6%) e tra coloro che hanno un titolo terziario di studio (74,1%)”.

Le statistiche sopra ricordate esprimono:

- Valori medi
- Valori mediani
- **Valori modali**
- Indici di dispersione
- Non so

Soluzione

È corretta la risposta n. 3: in riferimento alla popolazione che ha conseguito competenze digitali di base, il testo menziona le classi modali - considerate distintamente - per sesso, classe di età e titolo di studio.

L'Annuario Statistico Italiano, edizione 2023, p. 312, riporta le frequenze assolute, per classe di età, degli stranieri maschi con titolo di studio terziario, riferite al 2022. La tabella seguente ne mostra la distribuzione (valori espressi in migliaia di unità):

Classi di età	15-34	35-64	65-89
Frequenza assoluta	40	99	11

Rappresentando questi dati per mezzo di un istogramma, l'altezza del rettangolo relativo alla classe di età 35-64 è:

- 99
- 3,4
- 1,2
- 5,4
- Non so

Soluzione

È corretta la risposta n. 2: quando le ampiezze di classe non sono identiche viene meno la proporzionalità tra altezza del rettangolo e frequenza assoluta o relativa; dato che le frequenze rappresentano misure di superfici, ossia aree, le altezze corrispondono a densità, ovvero a rapporti tra frequenze e corrispondenti ampiezze di classe.

Nel caso preso in esame dobbiamo, quindi, dividere la frequenza (99, espressa in migliaia di unità) per l'ampiezza della classe (data dalla differenza tra 64 e 35, ovvero 29), ottenendo così, come risultato, 3,4, che sarà l'altezza del rettangolo relativo a quella classe di età.

L'Annuario Statistico Italiano, edizione 2023, p. 312, riporta le frequenze assolute, per classe di età, degli stranieri maschi con titolo di studio terziario, riferite al 2022. La tabella seguente ne mostra la distribuzione (valori espressi in migliaia di unità):

Classi di età	15-34	35-64	65-89
Frequenza assoluta	40	99	11

In quale classe di età si trova il terzo quartile Q_3 ?

- 15-34
- **35-64**
- 65-89
- In nessuna delle classi di età
- Non so

Soluzione

La somma delle frequenze è $n = 150$. La posizione del 3° quartile è la $n/4 \cdot 3$ -esima, ovvero la 112,5-esima osservazione, che cade nella classe d'età 35-64.

Nel capitolo dedicato all'obiettivo 3 (Salute e Benessere) del Rapporto SDGs 2024 dell'Istat, a pag. 53, si sottolinea che “se le persone meno istruite avessero avuto la stessa probabilità di morire delle persone più istruite, si sarebbero potute salvare circa 21.000 vite, cioè il 30% dei decessi prematuri avvenuti nel 2020 in Italia”.

L'affermazione indica che:

- In Italia, nel 2020, gli individui più istruiti avevano una probabilità più alta di morire rispetto a coloro che erano meno istruiti
- La probabilità di morire in Italia nel 2020 non dipendeva dal livello di istruzione
- **La probabilità di morire in Italia nel 2020 dipendeva dal livello di istruzione**
- In Italia, nel 2020, sono morte 21.000 persone
- Non so

Soluzione

L'affermazione sottolinea che in Italia nel 2020 si sarebbe potuto ridurre il numero di morti se gli individui meno istruiti avessero avuto la stessa probabilità di morire degli individui più istruiti. Purtroppo, la probabilità di morire varia condizionatamente al diverso livello di istruzione. Pertanto, è corretto affermare che la probabilità di morire in Italia nel 2020 dipendeva dal livello di istruzione ed era più bassa per le persone con un livello di istruzione più elevato, rispetto alle persone meno istruite.

Due caratteri quantitativi X e Y sono stati rilevati su un collettivo di 1.000 unità statistiche, osservando un coefficiente di correlazione lineare pari $-0,9$.

Quale di queste affermazioni è corretta?

- I due caratteri presentano un basso grado di dipendenza lineare
- Esiste una correlazione positiva tra i due caratteri
- Il carattere Y dipende linearmente dal carattere X
- **Tra i due caratteri esiste una relazione lineare decrescente quasi perfetta**
- Non so

Soluzione

La risposta corretta è la n. 4. Infatti, il coefficiente di correlazione lineare assume valori tra -1 e $+1$, con -1 in caso di relazione decrescente perfetta.

Al sig. Rossi è stato prescritto un nuovo farmaco e, leggendo il foglietto illustrativo, scopre che possono presentarsi, in maniera indipendente, due effetti collaterali indesiderati, A e B. Il primo (A) è molto comune ed è stato osservato in 1 paziente su 10. Il secondo (B) è meno comune e si presenta in 1 paziente su 100.

Qual è la probabilità che, nel caso del sig. Rossi, non si presenti nessuno dei due effetti indesiderati?

- 0,999
- **0,891**
- 0,001
- 0,11
- Non so

Soluzione

La risposta corretta è la n. 2.

Sia A l'evento "si presenta l'effetto indesiderato A" e B l'evento "si presenta l'effetto indesiderato B". Sappiamo che $P(A) = 1/10$ e $P(B) = 1/100$.

Considerando che i due effetti collaterali indesiderati si presentano in maniera indipendente, la probabilità richiesta è:

$$P(A^c \cap B^c) = P(A^c) * P(B^c) = (1 - 1/10) * (1 - 1/100) = 9/10 * 99/100 = 0,891$$