***Compito Statistica Sociale – tempo massimo: 1 ora e 30 minuti – Traccia D***

Nome: Cognome: Matricola:

Domande di teoria:

**Sezione n. 1**  *(si indichi se si ritengono vere o false le seguenti affermazioni. Per ogni risposta esatta, 1 punto)*

1. Un campione di dimensione = 20000 unità campionate, è un esempio di *Big Data*

V  F

1. Il coefficiente di variazione non può mai assumere il valore zero

V  F

1. Un diagramma a torta è particolarmente utile nel caso di variabili discrete o qualitative con poche modalità

V  F

1. Un indice di correlazione di Pearson pari a -0,95, indica una forte relazione inversa tra le due variabili

V  F

**Sezione n. 2**  *(Si scelga tra le seguenti la risposta corretta per completare la frase. Per ogni risposta esatta, 3 punti)*

1. Se la varianza in voti alla maturità di una scuola italiana (su scala 100) è superiore alla varianza in voti alla maturità di una scuola inglese (scala 20), allora…

I voti della scuola italiana sono più variabili

Occorre calcolare la deviazione standard per confrontarli

Occorre calcolare il coefficiente di variazione per confrontarli

Occorre calcolare la covarianza per confrontarli

1. “Abbiamo incrociato i dati relativi alle abitudini sul fumo (fumatore sì/no) e quelli relativi al genere (M/F) per quanto riguarda le quinte superiori della regione Friuli Venezia Giulia. Avendo ottenuto un indice **χ2** calcolato sulla tabella di contingenza pari a 0,54, possiamo affermare una sostanziale indipendenza tra le due variabili”. Colui che ha effettuato tale analisi …

Ha ragione

Dipende dalla natura dei dati, non si può affermare se ciò che dice sia vero o falso

Bisognerebbe conoscere il valore soglia (valore critico) associato alla tabella di contingenza

Il valore non è plausibile per un indice **χ2**

**Esercizio n. 1**  *(8 punti)*

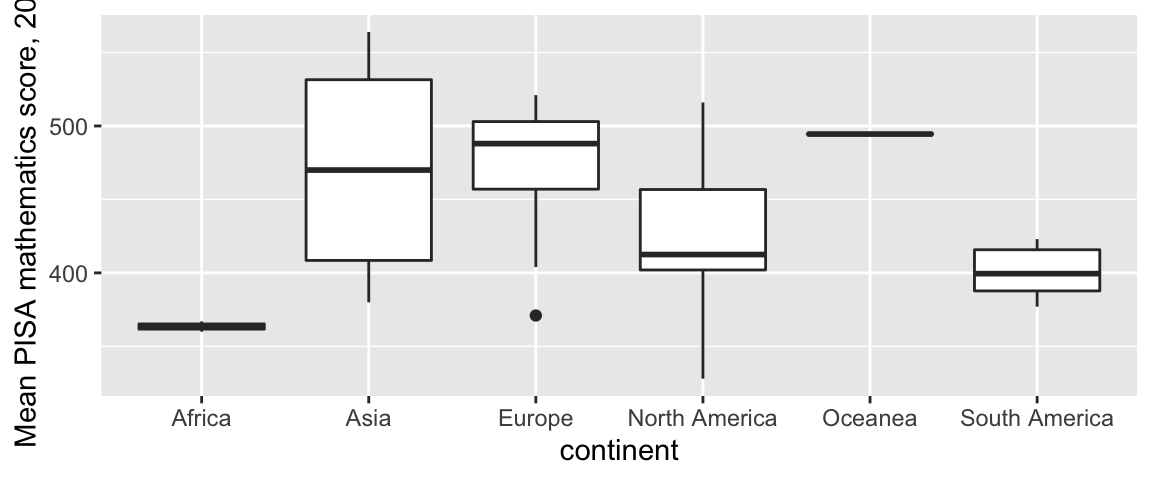
Uno studio condotto dall’Università di Trieste evidenzia come il numero di sospensioni scolastiche non sia uguale in tutti i mesi dell’anno accademico (dati 2017, scuole secondario di secondo grado del Comune di Trieste). La distribuzione è la seguente:

|  |  |
| --- | --- |
| ***Mese*** | ***N° Sospensioni*** |
| Gennaio | 44 |
| Febbraio | 35 |
| Marzo | 30 |
| Aprile | 28 |
| Maggio | 29 |
| Ottobre | 22 |
| Novembre | 25 |
| Dicembre | 42 |

Si indichi:

1. Media, moda, mediana e varianza del n° di sospensioni
2. Si suddivida la distribuzione in classi secondo un criterio e la si rappresenti graficamente

**Esercizio n. 2**  *(8 punti)*

**

1. Questo grafico è estratto da uno studio su dati di abilità in matematica (ogni osservazione rappresenta una Nazione) raggruppate poi per continente. Discutere che tipo di grafico è e la natura delle variabili che si ipotizza siano state utilizzate.
2. Discutere la presenza o meno di valori anomali (outlier) ed ipotizzare cosa rappresentino
3. Discutere in termini di mediana, di quartili e di variabilità il grafico

**Esercizio n. 3**  *(8 punti)*

Su 10 studenti triestini aspiranti educatori socio-pedagogici, sono state rilevate le ore di sonno in media a notte e le ore trascorse sui social (Facebook, Instagram ecc.) in media al giorno.

|  |  |
| --- | --- |
| ***Ore di sonno*** | ***Ore sui social*** |
| 4 | 4 |
| 6 | 4 |
| 6 | 4 |
| 7 | 3 |
| 7 | 2 |
| 7 | 2 |
| 7 | 4 |
| 8 | 2 |
| 9 | 1 |
| 10 | 1 |

1. Si adotti una adeguata rappresentazione grafica per rappresentare contemporaneamente le due variabili
2. Si evidenzi la natura della relazione tra le due variabili e si interpreti in risultato (avvalendosi, se utile, anche di un indice oltre che della rappresentazione grafica)