

Metodi Matematici e Informatici per la Biologia
a.a. 2023-2024
(Agliari - Panati - Presilla - Simonella)
18 giugno 2024
Prova di Verifica I
Soluzioni

punteggio massimo $33 = 7 + 6 + 6 + 8 + 6$

Nome:

Cognome:

Matricola:

Esercizio 1 **7 punti**

I dati seguenti indicano i prezzi di un chilo di pane in euro al kg, rilevati nel 2023 in 30 diverse panetterie di una metropoli europea:

4.78, 4.73, 5.22, 4.46, 3.79, 4.37, 4.71, 3.63, 4.16, 3.97, 4.74, 3.64, 3.93, 4.06, 3.63, 4.7, 3.92, 4.87, 4.3, 3.78, 4.78, 4.12, 3.85, 4.6, 4.01, 4.62, 4.17, 4.63, 3.74, 4.49.

- 2** **1.1** Si calcolino media, moda, mediana e varianza campionarie, ricordandone le definizioni.

La media campionaria è 4.28€; la mediana campionaria è 4.235€, la varianza campionaria è 0.53€².
0.195

- 1** **1.2** Si spieghi come cambia la mediana se si aggiunge al campione il dato 4.20€.

La nuova mediana è 4.20€.

- 1** **1.3** Si disegni il boxplot del campione di dati.

Primo e terzo quartile 3.92, 4.7€, valori min/max 3.63, 5.22€.

- 1** **1.4** Si indichi la proporzione di dati compresa tra i valori $\bar{X} \pm 2S$ e la si confronti con quella di una distribuzione approssimativamente normale.

*I valori estremi sono **3.39**, **5.16**€. I dati compresi tra tali valori sono il **96.6%** dei dati complessivi. Per una distribuzione normale ci si aspetta che il 95% dei dati sia nell'intervallo.*

- 2** **1.5** Si calcolino, illustrando il procedimento:

(a) il 40-esimo percentile del campione di dati;

$n \times p = 30 \times 0.40 = 12$, da cui si trova il valore $(4.06 + 4.12)/2 = 4.09€$.

(b) il 95-esimo percentile del campione di dati.

$n \times p = 30 \times 0.95 = 28.5$, da cui si trova il dato 4.87€.

Esercizio 2 6 punti

Un corso intensivo di specializzazione molto frequentato si svolge ogni trimestre. Tenendo conto dei risultati su diversi trimestri, si osserva che il numero medio di partecipanti che supera il test finale con il massimo punteggio è 3.5 a trimestre. Si indichi una ragionevole distribuzione di probabilità per i partecipanti che, ogni trimestre, completano il corso con il massimo punteggio.

2 *Distribuzione di Poisson di parametro $\lambda = 3.5$.*

Utilizzando tale distribuzione, si calcoli la probabilità che nel prossimo trimestre:

2 (a) nessun partecipante ottenga il massimo dei voti;

0.03.

2 (b) almeno 3 partecipanti ottengano il massimo dei voti.

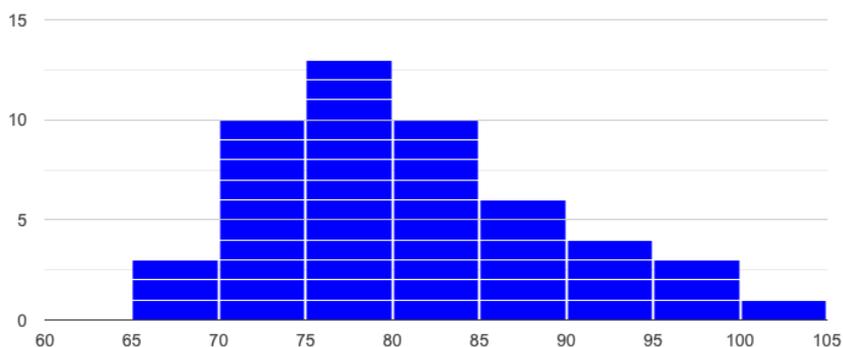
$1 - 0.03(1 + 3.5 + (3.5)^2/2) = 0.68$.

Esercizio 3 6 punti

In un'azienda manifatturiera, vengono misurati i tempi di produzione di un bene, per 50 sessioni di lavoro. Si ottengono i dati seguenti, espressi in minuti:

76, 86, 83, 96, 73, 75, 67, 80, 79, 77, 79, 69, 86, 70, 75, 79, 81, 77, 94, 83, 91, 74, 72, 92, 90, 75, 72, 86, 85, 83, 71, 88, 78, 98, 69, 79, 77, 70, 78, 72, 84, 100, 83, 83, 82, 85, 72, 83, 98, 72.

2 **3.1** Si dividano i dati in 8 classi di ampiezza pari a 5 minuti, a partire dal valore di 65 minuti (non presente nei dati). Si riportino, per ogni classe, le frequenze relative e il corrispondente istogramma.



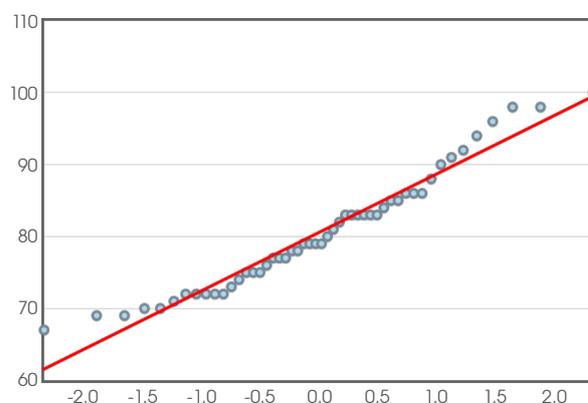
classe (minuti)	freq. ^{assoluta} rel.
[65,70)	3
[70,75)	10
[75,80)	13
80-84	10
85-89	6
90-94	4
95-99	3
[100,105)	1

- 2 **3.2** Si determinino degli stimatori per valore atteso e deviazione standard della distribuzione dei dati in considerazione, a partire dai dati.

La media campionaria è 80.5 minuti; la deviazione standard campionaria è 8.3 minuti.²

- 2 **3.3** Si dica se è o meno ragionevole che i dati provengano da una distribuzione normale, motivando la risposta.

Plot dei quantili dei dati versus quantili normali standard: ipotesi di normalità ragionevole.

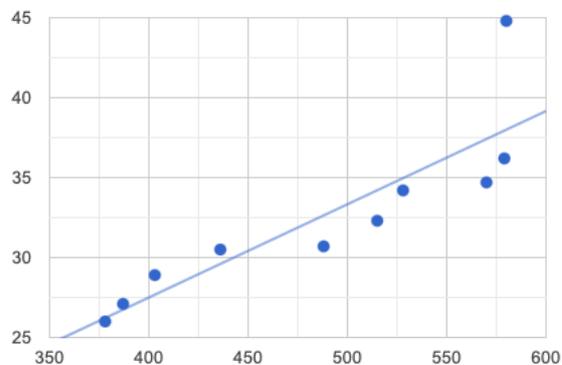


Esercizio 4 8 punti

I seguenti dati indicano la resa stagionale di patate in tonnellate per ettaro (variabile y), in dieci diverse località, tutte aventi la stessa qualità di terreno. I dati sono messi in relazione con la quantità stagionale di precipitazioni in millimetri (variabile x).

- 2 **4.1** Si osservi il grafico di dispersione e si calcoli l'equazione della retta di regressione con il metodo dei minimi quadrati, dopo aver eliminato eventuali outliers. Si specifichino le unità di misura dei parametri della retta.

x	y
579	36.2
378	26.0
436	30.5
528	34.2
488	30.7
403	28.9
515	32.3
387	27.1
580	44.8
570	34.7



Possiamo rimuovere il dato (580, 44.8).

[ATTENZIONE: DATO RIMOSSO DAL TESTO NELLA VERSIONE DEFINITIVA]

Si ottiene $Y = 10.4 + 0.0437X$. 10.4 t/ha 0.0437 t/(ha mm)

- 2 4.2 Si calcoli il coefficiente di determinazione della retta di regressione ottenuta e se ne spieghi l'interpretazione.

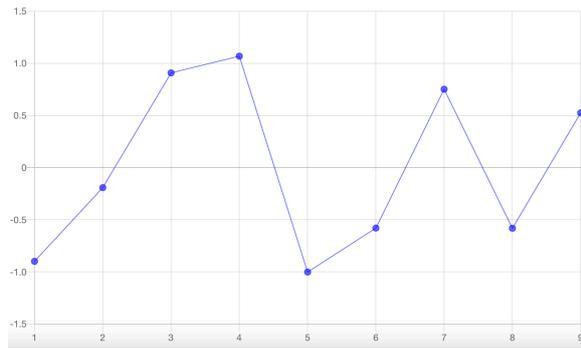
$$R^2 = 0.945.$$

- 2 4.4 Si usi il modello di regressione per predire la resa stagionale in presenza di = 470 mm di precipitazioni.

30.9 ^{t/ha}mm.

1 ettaro = 100 are (da cui il simbolo ha) 1 ara = 100 m²
 ara = unità di misura di superficie usata nel Catasto Terreni

- 2 4.3 Analizzando i residui, si dica se è o meno ragionevole che il modello lineare sia un buon modello per rappresentare la dipendenza di y da x .



Buon modello

Esercizio 5 6 punti

Nella seguente tabella è indicato il numero di soggetti sopravvissuti a seguito di esposizione a radiazione in una esperienza di laboratorio, effettuata su due diverse specie animali. Vengono utilizzati rispettivamente 72 soggetti per la specie A e 118 soggetti per la specie B.

Sopravvivenza	Specie A	Specie B
Sì	66	88
No	6	30

Si verifichi l'ipotesi nulla secondo cui la probabilità di sopravvivenza è indipendente dalla specie (A o B), fissando il livello di significatività del test al 1%.

La tabella delle frequenze attese è

Sopravvivenza	Specie A	Specie B
Sì	58.36	95.64
No	13.64	22.36

da cui si ricava la statistica di test

$$ST = \frac{(66 - 58.36)^2}{58.36} + \frac{(88 - 95.64)^2}{95.64} + \frac{(6 - 13.64)^2}{13.64} + \frac{(30 - 22.36)^2}{22.36}$$

4 ovvero $ST \simeq 8.5$. La variabile chi-quadro a 1 grado di libertà corrispondente a un livello di fiducia dello 0.01 ha valore 6.6. L'ipotesi nulla è rifiutata. 2