

Metodi Matematici e Informatici per la Biologia
Appello del 16 Giugno 2016

Nome: _____ Cognome: _____

Matricola: _____

Specificate se su Infostud dovete verbalizzare:

a) 9 + 3 crediti Nuovo Ordinamento (Calcolo, Biostatistica+MMIB)

oppure

b) SOLO MMIB 3 crediti Vecchio Ordinamento

ESERCIZIO 1

In un test sull'utilizzo del mouse vengono coinvolti 20 individui. Ogni individuo viene posto di fronte ad uno schermo su cui appaiono diverse figure; il soggetto deve cliccare sul mouse il più presto possibile non appena vede una determinata figura sullo schermo.

La tabella sottostante riporta i tempi di reazione registrati.

Tempo di reazione (ms)

115

96

110

100

111

101

111

106

96

96

95

96

96

106

100

113

122

111

95

108

i) Per questi dati costruite un opportuno istogramma riportando, oltre alla figura, anche una descrizione delle classi considerate.

ii) Specificate almeno un altro tipo di grafico che può essere usato per descrivere questi dati.

iii) Calcolate la media e la deviazione standard per i dati di questo campione.

iv) Questi dati si distribuiscono approssimativamente secondo una normale? Giustificate la vostra risposta.

v) Supponete che questo campione di dati provenga da una distribuzione approssimativamente normale con media e deviazione standard pari a quelle calcolate sul campione (punto iii). Qual è la probabilità che venga registrato un tempo di reazione superiore a 120 ms?

ESERCIZIO 2

Quando è stata introdotta la penicillina, nel 1944, più del 94% dei ceppi di *Staphylococcus aureus* isolati era sensibile; nel 1987, il 20% dei ceppi di *S. aureus* isolati era resistente alla penicillina; nel 2007 questa percentuale è salita circa all'80%.

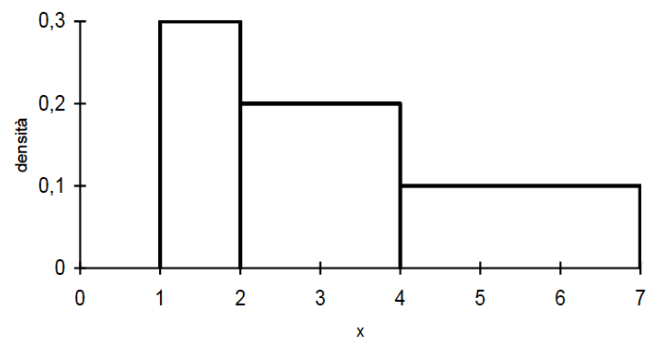
i) Se si vuole stimare la probabilità P_{116} che in una coltura di 6 ceppi, almeno uno sia resistente, quale modello si potrebbe usare? Sotto quali condizioni?

ii) Calcolare di quanto è aumentata P_{116} tra il 1987 e il 2007.

$$P_{116}(2007) / P_{116}(1987) = \dots$$

ESERCIZIO 3

Calcolare la probabilità che la variabile aleatoria descritta dall'istogramma sottostante assuma un valore compreso tra 4 e 7.



ESERCIZIO 4

i) In una distribuzione normale il 92mo e il 96mo percentile sono indicati con x_{92} e x_{96} , rispettivamente. Analogamente, si indicano con x_{43} e x_{47} il 43mo percentile ed il 47mo percentile, rispettivamente. Indicare quale/quale delle seguenti affermazioni è/sono corretta/corrette

- a- $(x_{96} - x_{92})$ è circa uguale a $(x_{47} - x_{43})$
- b- $(x_{96} - x_{92})$ è maggiore di $(x_{47} - x_{43})$
- c- $(x_{96} - x_{92})$ è minore di $(x_{47} - x_{43})$
- d - In generale, nessuna delle precedenti è corretta

ii) Se la distribuzione fosse uniforme su intervallo $[0, 100]$, quale/quale delle seguenti affermazioni è/sono corretta/corrette

- a- $(x_{96} - x_{92})$ è circa uguale a $(x_{47} - x_{43})$
- b- $(x_{96} - x_{92})$ è maggiore di $(x_{47} - x_{43})$
- c- $(x_{96} - x_{92})$ è minore di $(x_{47} - x_{43})$
- d - In generale, nessuna delle precedenti è corretta

ESERCIZIO 5

Si vuole testare l'efficacia del farmaco idrossiurea per ridurre il dolore nelle persone affette da anemia falciforme. Vengono individuati 300 soggetti malati per partecipare alla ricerca; tali soggetti presentano, ai fini della ricerca, analoghe caratteristiche, ovvero costituiscono un campione omogeneo. Si individuano due gruppi: al primo, formato da 150 soggetti, è stato dato il farmaco, all'altro, sempre di 150 soggetti, una sostanza placebo.

- i) Si tratta di uno studio di osservazione o di un esperimento? Perché?
- ii) Come divideresti i 300 soggetto nei due gruppi?
- iii) Indica i soggetti, il trattamento e la variabile di risposta.

ESERCIZIO 6

In una ricerca (Lahvis et al., 1995) sono stati eseguiti test per analizzare gli effetti dell'esposizione a policlorobifenili (PCB) nei delfini dal naso a bottiglia (*Tursiops truncatus*).

Per sei delfini sono stati raccolti dati sui livelli di PCBs nel sangue e si è studiato se le concentrazioni di PCB aumentano in funzione dell'età dell'animale.

I dati sono mostrati nella tabella sottostante

<i>Età (anni)</i>	<i>Concentrazione di PBC (ng/g)</i>
1	1.5
9	9.1
13	5.8
15	10.7
21	17.6
28	13.6
32	19.1

Con gli strumenti a vostra disposizione analizzate i dati per capire se esiste una relazione tra concentrazioni di PCB ed età dell'animale, riportando i passaggi, le misure ed i grafici principali sul foglio.

Metodi Matematici e Informatici per la Biologia
Appello del 16 Giugno 2016
CORREZIONI

ESERCIZIO 1

In un test sull'utilizzo del mouse vengono coinvolti 20 individui. Ogni individuo viene posto di fronte ad uno schermo su cui appaiono diverse figure; il soggetto deve cliccare sul mouse il più presto possibile non appena vede una determinata figura sullo schermo.

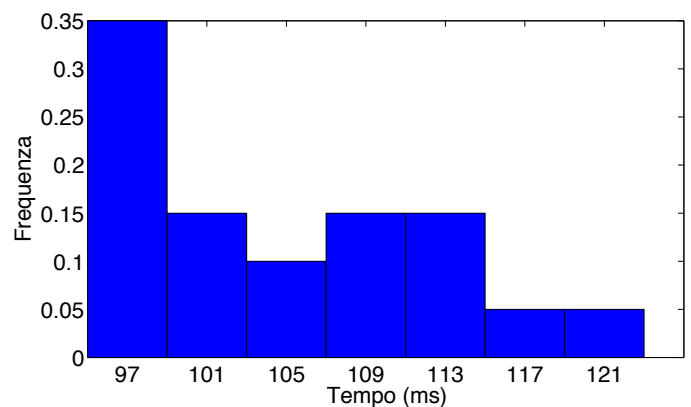
La tabella sottostante riporta i tempi di reazione registrati.

Tempo di reazione (ms)

115
96
110
100
111
101
111
106
96
96
95
96
96
106
100
113
122
111
95
108

- i) Per questi dati costruite un opportuno istogramma riportando, oltre alla figura, anche una descrizione delle classi considerate.

Classe	Count	Frequenza
[95-99)	7	0.35
[99,103)	3	0.15
[103,107)	2	0.10
[107,111)	3	0.15
[111,115)	3	0.15
[115,119)	1	0.05
[119,123)	1	0.05



- ii) Specificate almeno un altro tipo di grafico che può essere usato per descrivere questi dati.

Diagramma ramo-foglia

- iii) Calcolate la media e la deviazione standard per i dati di questo campione.

Media = 104.2500 ms; Deviazione Standard = 8.2454 ms

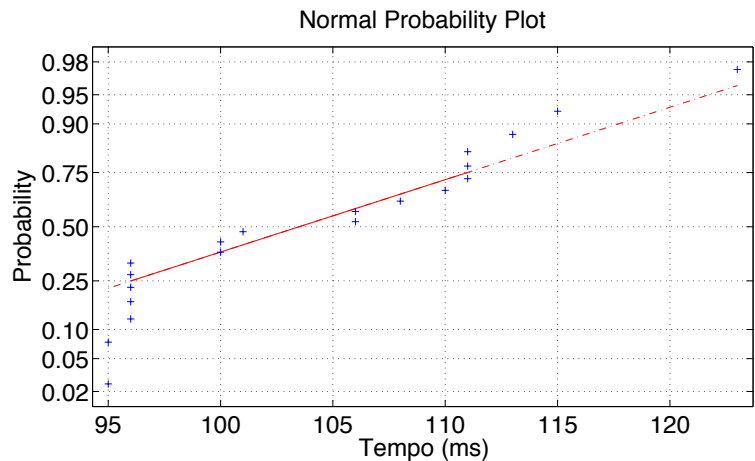
iv) Questi dati si distribuiscono approssimativamente secondo una normale? Giustificate la vostra risposta.

L'istogramma del punto precedente mostra asimmetria a destra per cui, da una prima analisi qualitativa, il modello normale non sembra un buon modello.

Curtosi = -0.7356 -> distribuzione platicurtica (più piatta di una normale)

Indice di asimmetria = 0.4914 -> positivo -> asimmetria a destra

Il normal plot mostra un andamento che, per tempi brevi, non rispetta l'andamento lineare. La forma del normal plot suggerisce che la distribuzione dei tempi sia asimmetrica a destra



Tutte le verifiche riportate qui sopra concordano col fatto che il modello normale non sia un buon modello.

v) Supponete che questo campione di dati provenga da una distribuzione approssimativamente normale con media e deviazione standard pari a quelle calcolate sul campione (punto iii). Qual è la probabilità che venga registrato un tempo di reazione superiore a 120 ms?

$$\mu = 104.2500 \text{ ms}$$

$$\sigma = 8.2454 \text{ ms}$$

$$\text{Standardizzo} \rightarrow z = (x - \mu) / \sigma = (120 - 104.25) / 8.2454 = 1.9102$$

$$\text{Tabelle} \rightarrow P(z < 1.91) \approx 0.97$$

$$P(x > 120) \approx P(z > 1.91) = 1 - P(z < 1.91) \approx 0.03 \rightarrow 3\%$$

ESERCIZIO 2

Quando è stata introdotta la penicillina, nel 1944, più del 94% dei ceppi di Staphylococcus aureus isolati era sensibile; nel 1987, il 20% dei ceppi di S. aureus isolati era resistente alla penicillina; nel 2007 questa percentuale è salita circa all'80%.

i) Se si vuole stimare la probabilità P_{116} che in una coltura di 6 ceppi, almeno uno sia resistente, quale modello si potrebbe usare? Sotto quali condizioni?

Posso usare un modello binomiale sotto le seguenti condizioni:

- La probabilità di resistenza p sia la stessa per tutti i ceppi
- Le probabilità di resistenza dei vari ceppi siano indipendenti tra loro
- Esiste un numero massimo N di ceppi che possono mostrare resistenza

In questo caso:

$$P(k) = \text{Binom}(N, k) p^k (1-p)^{N-k} \rightarrow \text{probabilità che } k \text{ siano resistenti}$$

$$P(1) + P(2) + \dots + P(N) = 1 - P(0) \rightarrow \text{probabilità che almeno uno sia resistente}$$

$$P_{1N} = 1 - P(0) = 1 - (1-p)^N$$

ii) Calcolare di quanto è aumentata P_{116} tra il 1987 e il 2007.

$$p_{2007} = 0.80$$

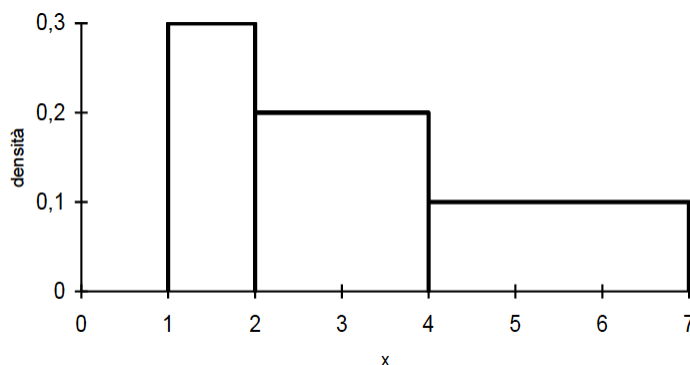
$$p_{1987} = 0.20$$

$$P_{116}(2007) / P_{116}(1987) = [1 -(1-p_{2007})^N] / [1 -(1-p_{1987})^N] = [1 - 0.2^6] / [1 - 0.8^6] \approx 0.9999 / 0.7379 \approx 1.355$$

ESERCIZIO 3

Calcolare la probabilità che la variabile aleatoria descritta dall'istogramma sottostante assuma un valore compreso tra 4 e 7.

$$P(4 < x < 7) = 0.3$$



ESERCIZIO 4

i) In una distribuzione normale il 92mo e il 96mo percentile sono indicati con x_{92} e x_{96} , rispettivamente. Analogamente, si indicano con x_{43} e x_{47} il 43mo percentile ed il 47mo percentile, rispettivamente. Indicare quale/quali delle seguenti affermazioni è/sono corretta/corrette

- a- $(x_{96} - x_{92})$ è circa uguale a $(x_{47} - x_{43})$
- b- $(x_{96} - x_{92})$ è maggiore di $(x_{47} - x_{43})$
- c- $(x_{96} - x_{92})$ è minore di $(x_{47} - x_{43})$
- d - In generale, nessuna delle precedenti è corretta

ii) Se la distribuzione fosse uniforme su intervallo $[0, 100]$, quale/quali delle seguenti affermazioni è/sono corretta/corrette

- a- $(x_{96} - x_{92})$ è circa uguale a $(x_{47} - x_{43})$
- b- $(x_{96} - x_{92})$ è maggiore di $(x_{47} - x_{43})$
- c- $(x_{96} - x_{92})$ è minore di $(x_{47} - x_{43})$
- d - In generale, nessuna delle precedenti è corretta

ESERCIZIO 5

Si vuole testare l'efficacia del farmaco idrossiurea per ridurre il dolore nelle persone affette da anemia falciforme. Vengono individuati 300 soggetti malati per partecipare alla ricerca; tali soggetti presentano, ai fini della ricerca, analoghe caratteristiche, ovvero costituiscono un campione omogeneo. Si individuano due gruppi: al primo, formato da 150 soggetti, è stato dato il farmaco, all'altro, sempre di 150 soggetti, una sostanza placebo.

i) Si tratta di uno studio di osservazione o di un esperimento? Perché?

Esperimento perché i soggetti vengono sottoposti ad un trattamento

ii) Come divideresti i 300 nei due gruppi?

Associo ad ogni individuo un numero da 1 a 300 ed estraggo in maniera casuale 150 numeri interi tra 1 e 300. Gli individui corrispondenti ai numeri estratti costituiranno uno dei due gruppi.

iii) Indica i soggetti, il trattamento e la variabile di risposta.

Soggetto = Malato di anemia falciforme

Trattamento = idrossiurea

Variabile di risposta = Intensità e frequenza dei dolori

ESERCIZIO 6

In una ricerca (Lahvis et al., 1995) sono stati eseguiti test per analizzare gli effetti dell'esposizione a policlorobifenili (PCB) nei delfini dal naso a bottiglia (*Tursiops truncatus*).

Per sei delfini sono stati raccolti dati sui livelli di PCBs nel sangue e si è studiato se le concentrazioni di PCB aumentano in funzione dell'età dell'animale.

I dati sono mostrati nella tabella sottostante

Età (in anni)	Concentrazione di PBC (ng/g)
1	1.5
9	9.1
13	5.8
15	10.7
21	17.6
28	13.6
32	19.1

Con gli strumenti a vostra disposizione analizzate i dati per capire se esiste una relazione tra concentrazioni di PCB ed età dell'animale, riportando i passaggi, le misure ed i grafici principali sul foglio.

Diagramma di dispersione

-> Relazione positiva

Un modello lineare sembra ragionevole

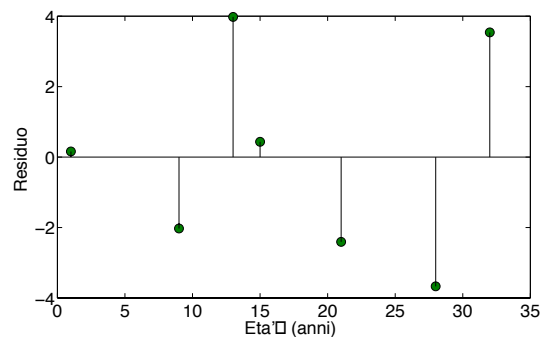
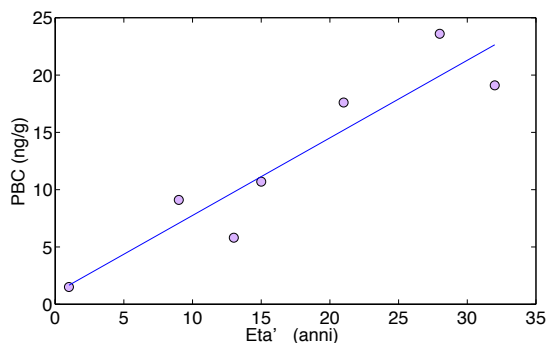
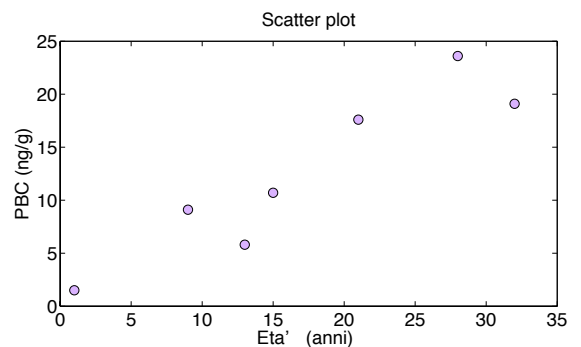
Quantifico attraverso il coefficiente di Pearson

$r = 0.79$ -> il 62% della variabilità è spiegata dal modello lineare

Retta di regressione $y = a x + b$

$a = 0.6768$, $b = 0.9804$

Il grafico dei residui mostra che i residui sono distribuiti in maniera abbastanza simmetrica rispetto allo zero, ovvero non ci sono andamenti sistematici.



PUNTEGGI

ESERCIZIO 1

- i) 2.5
- ii) 1
- iii) 1
- iv) 1+2+3
- v) 2.5

ESERCIZIO 2

- i) 2.5
- ii) 2

ESERCIZIO 3

1.5

ESERCIZIO 4

- i) 2
- ii) 2

ESERCIZIO 5

- i) 1
- ii) 1
- iii) 2

ESERCIZIO 6

1 + 2 + 3