

Metodi Matematici e Informatici per la Biologia
Appello del 5- 7- 2018

Nome: _____ Cognome: _____

Matricola: _____

ESERCIZIO 1

Su 90 studenti dell'ultimo anno di un istituto superiore viene eseguita la simulazione di un test di ammissione all'università. I risultati del test (in una scala di punteggi da 0 a 60) vengono raggruppati in classi secondo la seguente tabella:

(0-10]; (10-20]; (20-30]; (30-40]; (40-50]; (50-60]

6 12 22 19 18 13

dove nella prima riga sono riportati gli intervalli dei punteggi e sotto le frequenze assolute.

Mediante un test del chi quadro dire se è accettabile l'ipotesi nulla (e con quale livello di significatività) che i punteggi siano uniformemente distribuiti nel range 0-60 (riportando i passaggi).

Soluzione.

Sotto l'ipotesi che i punteggi siano uniformemente distribuiti nel range 0-60 la probabilità di aver un punteggio in una delle classi indicate è $1/6$, per cui le frequenze teoriche sono tutte uguali e pari a $90/6=15$

tabella di contingenza delle frequenze teoriche

(0-10]; (10-20]; (20-30]; (30-40]; (40-50]; (50-60]

15 15 15 15 15 15

$\chi^2=(81+9+49+16+9+4)/15=11,2$. Va confrontato con il χ^2 a 5 gradi di libertà. Scelgo un livello di significatività al 5%

Osservo che $11,2 > 11,07$, che corrisponde ad un livello di significatività del 5%, quindi l'ipotesi va scartata perchè il livello di significatività è inferiore al 5%.

ESERCIZIO 2

In un esperimento per valutare se l'insegnamento della statistica basato sull'approccio dell'apprendimento per problemi è migliore dell'insegnamento tradizionale, metà degli studenti di uno dei corsi di statistica della facoltà di medicina sono stati assegnati casualmente a un tipo di insegnamento e metà all'altro.

i) dite di quale tipo di esperimento si tratta

Esperimento completamente randomizzato

ii) spiegate qual è la procedura per l'assegnazione casuale degli studenti ai 2 tipi di corso

Numero gli studenti da 1 a N (=numero totale degli studenti del corso), poi estraggo a caso (senza ripetizione) $N/2$ numeri da un insieme di numeri $\{1, \dots, N\}$. Gli studenti corrispondenti ai numeri estratti formeranno il primo gruppo, gli altri il secondo.

iii) qual è la popolazione di riferimento in questo esperimento?

Studenti di statistica della facoltà di medicina

iv) in questo studio, sono state misurate le variabili riportate sotto. Per ciascuna, dite di che tipo di variabile si tratta e qual è il miglior indice di tendenza centrale e di dispersione da usare per quella variabile:

a) Tempo per completare l'esame. Non c'era limite di tempo (alcuni studenti hanno terminato tardi)

Variabile quantitativa continua; media, varianza.

b) Il tipo di mal di testa che può cogliere gli studenti all'esame (emicrania a grappolo o tensiva)

Variabile qualitativa categorica; moda.

c) Per i laureati in medicina che hanno proseguito nella ricerca, il numero di lavori che sono stati rifiutati dalle riviste perché l'analisi statistica era inappropriata.

Variabile quantitativa discreta; mediana, distanza interquartile.

ESERCIZIO 3

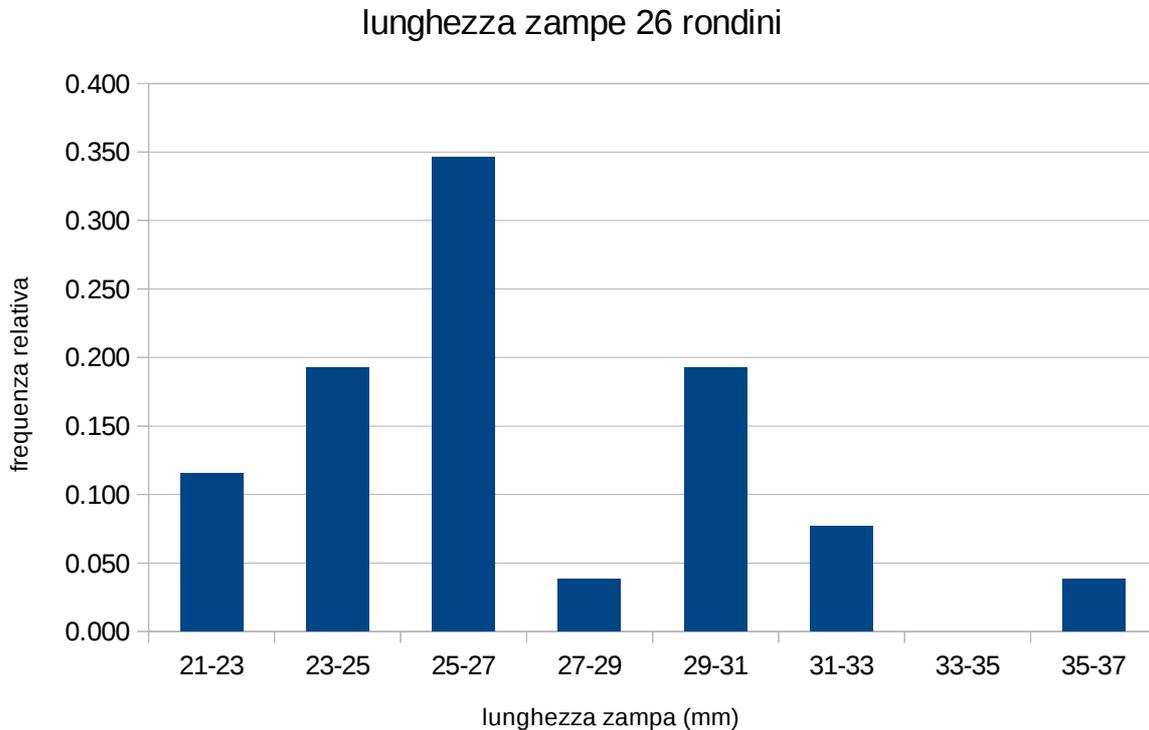
I dati che seguono riportano le misure in mm delle lunghezze di una zampa di 26 rondini.

22.0	24.0
22.0	25.2
23.5	25.2
23.5	25.3
24.0	25.3
21.5	25.4
30.0	26.0
30.0	26.0
30.0	26.0
32.0	26.0
32.0	28.0
35.1	29.2
24.0	29.1

i) Per questi dati, considerate 8 classi di ampiezza pari a 2 a partire dal valore 21 (valore non compreso nei dati) e calcolate le frequenze relative per ogni classe, riportandole qui sotto.

Etichetta classe	frequenza relativa
21-23	0.115
23-25	0.192
25-27	0.346
27-29	0.038
29-31	0.192
31-33	0.077
33-35	0
35-37	0.038

ii) Poi disegnate l'istogramma, riportando il grafico sul foglio qui sotto. Specificate cosa riportate sugli assi



iii) per questi dati disegnate il normal plot senza riportarlo sul foglio. Osservando il grafico, si può dire che i dati si distribuiscono approssimativamente secondo una distribuzione normale? Spiegate perché si o perché no.

I dati si distribuiscono approssimativamente secondo una distribuzione normale; dal normal plot si evince un andamento lineare tra osservazioni e quantili normali standard, confermato da un coefficiente di determinazione alto ($r^2 = 0.94$). Inoltre l'indice di asimmetria è basso (0.71), come anche l'indice di curtosi (-0.02).

iv) dite cosa si riporta sui 2 assi di un normal plot

Asse X: osservazioni; asse Y: quantili normali standardizzati

ESERCIZIO 4

L'altezza della popolazione francese si distribuisce secondo una normale con media $\mu = 166$ cm e scarto standard $\sigma = 10$ cm.

i) La probabilità che un francese scelto a caso superi 160 cm e' maggiore della probabilità che sia piu' basso di 180 cm.

Usando le tavole e riportando i calcoli, dite se è vero o falso.

Indico con X la v.a. di media μ e deviazione standard σ , e con $Z = (X - \mu) / \sigma$ la v.a. normale standard.
 $P(X > 160) = P(Z > (160 - \mu) / \sigma) = P(Z > -0.6)$. Per simmetria $P(Z > -0.6) = P(Z < 0.6) = 0.7257$.
 $P(X < 180) = P(Z < (180 - \mu) / \sigma) = P(Z < 1.4) = 0.9192$
 Quindi l'affermazione $P(X > 160) > P(X < 180)$ è FALSA

ii) Sempre, usando le tavole e riportando i calcoli spiegate se è vero che piu' del 50% della popolazione francese ha altezze comprese tra 155 e 175 cm.

$$P(155 < X < 175) = P(-1.1 < Z < 0.9) = P(Z < 0.9) - P(Z < -1.1) = 0.8159 - (1 - 0.8643) = 0.6802.$$

VERO: circa il 68% della popolazione ha un'altezza compresa tra 155 e 175 cm.

ESERCIZIO 5

Alcuni tipi di cellule in un tessuto rappresentano solo una piccola parte di tutta la popolazione di cellule, come ad es. le cellule staminali ematopoietiche nel sangue. Supponete che un particolare tipo di cellule rare sia presente, mediamente, nel rapporto di 1 cellula su 100.000.

i) Volendo calcolare la probabilità di non trovare nessuna di queste cellule rare in una piccola porzione di tessuto, quale distribuzione di probabilità usereste?

Distribuzione di Poisson.

ii) Dopo averle verificate, riportate le condizioni che sono necessarie per l'applicabilità della distribuzione che avete scelto.

La probabilità stimata di trovare una cellula rara è $p=1/100000$. Se le cellule presenti nel tessuto sono $N \gg 1$, con N tale che Np è circa ordine 1-10, allora posso usare la distribuzione di Poisson di parametro $\lambda=pN$.

ESERCIZIO 6

Considerate queste coppie di dati.

<u>x</u>	<u>y</u>
80.7	2.00
71.8	3.64
75.6	3.42
72.7	2.84
75.3	1.83
73.6	2.82
75.1	4.04
74.0	2.56
73.8	0.80
70.4	4.19

Dopo aver osservato il grafico di dispersione e calcolato l'equazione della retta di regressione

i) dite a quale valore x corrisponde il residuo maggiore in valore assoluto

$x=73.8$

ii) il punto corrispondente a tale valore x può essere considerato un outlier o un'osservazione influente rispetto alla retta di regressione? **Spiegate la vostra scelta**

È un outlier

iii) quale percentuale di variabilità delle y è spiegata dal modello di regressione dei minimi quadrati?

16.6%

