

1) I pesi di una data specie di uccelli in un'area boschiva seguono una distribuzione normale di media 1500 grammi e deviazione standard 500 grammi. Calcolare quanto vale la probabilità di avere uccelli dal peso compreso tra 1500 grammi e 2000 grammi.

$$P(1500 < x < 2000) = P(0 < z < 1) = P(z < 1) - P(z < 0) = 0,34$$

2) In un concorso pubblico i risultati di una prova scritta seguono una distribuzione normale con media 80 e deviazione standard 20. Sapendo che il punteggio minimo per accedere all'orale è 63, determinare la percentuale di candidati esclusa dalla prova orale.

$$P(x < 63) = P(z < z^*)$$

$$\text{con } z^* = (63 - 80) / 20 = -0,85$$

$$P(z < -0,85) = P(z > 0,85) = 1 - P(z < 0,85) = 0,20$$

3) Il numero di messaggi spam ricevuti in una settimana è mediamente 50. Con quale probabilità passerà un giorno senza ricevere messaggi spam?

Si tratta di una distribuzione di Poisson. Il numero medio di messaggi spam giornalieri sarà  $50/7 = 7,14$

$$\text{da cui } P(0) = e^{-7,14} = 8 \cdot 10^{-4}$$

4) In una famiglia con 5 figli determinare la probabilità che tutti abbiano lo stesso sesso (supporre  $p(\text{maschio}) = p(\text{femmina}) = 1/2$ ).

Si adopera una distribuzione binomiale.

Chiaramente la probabilità che i 5 figli siano tutti maschi è  $(1/2)^5$

Anche la probabilità che i 5 figli siano tutti femmina è  $(1/2)^5$ , e la probabilità cercata è la somma delle due.

5) Un ricercatore usa un modello di regressione semplice per fare previsioni sul costo  $y$  del riscaldamento mensile (euro) sulla base della dimensione  $x$  di un appartamento (metri quadrati). La correlazione tra il costo previsto e la dimensione di una casa è 0,70.

Qual è l'interpretazione corretta di questo risultato?

il 49% della variabilità del costo del riscaldamento può essere spiegata dalla dimensione della casa.

6) Riportate i seguenti dati in un diagramma xy, riportando anche sullo stesso grafico la retta di regressione dei minimi quadrati. Spiegate se il modello lineare rappresenta un buon modello per tali dati.

| <u>x</u> | <u>y</u> |
|----------|----------|
| 10       | 7.46     |
| 8        | 6.77     |
| 13       | 12.74    |
| 9        | 7.11     |
| 11       | 7.81     |
| 14       | 8.84     |
| 6        | 6.08     |
| 4        | 5.39     |
| 12       | 8.15     |
| 7        | 6.42     |
| 5        | 5.73     |

La regressione lineare non è un buon modello per questi dati. In particolare la coppia di valori (13; 12,74) si discosta dal trend, il valore di  $R^2 = 0,68$  non è vicino a 1 e il grafico dei residui non presenta un andamento casuale.

7) Del residuo di una osservazione influente possiamo dire che:  
è molto distante dagli altri residui in direzione orizzontale

8) Se in un grafico dei residui questi ultimi sono piccoli in valore assoluto cosa possiamo concludere?  
nulla, dipende se sono distribuiti casualmente oppure no

9) In una classe di 15 allievi si sono conseguiti i seguenti risultati delle prove di algebra e di fisica:

algebra: 9 5 6 6 6 7 7 7 8 8 8 8 9 9 10

fisica: 3 3 5 6 6 6 5 6 6 8 9 8 8 9 9

calcolate la retta di regressione dei minimi quadrati dove algebra è la variabile esplicativa x e fisica è la variabile di risposta y

Quale delle seguenti è l'equazione della retta?

$$y = 0,147 + 0,839x$$

10) Uno studio statistico sulle alluvioni che si sono verificate nei mesi estivi (giugno, luglio, agosto) in una certa regione tropicale dal 2014 al 2018 indica che il 15 % delle alluvioni si sono verificate a giugno, il 25 % a luglio e il 60 % ad agosto. I dati sulle alluvioni nel 2019 indicano che nel mese di giugno si sono verificate 28 alluvioni, 54 a luglio e 78 ad agosto. Prendendo come ipotesi nulla che i dati del 2019 seguano la stessa distribuzione dello studio riportato

Posso rifiutare l'ipotesi nulla a un livello del 5%

11) Uno studio sui danni del fumo passivo in gravidanza sostiene che i bambini esposti al fumo di sigaretta dei loro padri mentre sono ancora in grembo sono più propensi a sviluppare l'asma. La seguente tabella di contingenza riporta i dati relativi a un campione di 756 bambini di sei anni

|                    | Asma | No asma | Totale |
|--------------------|------|---------|--------|
| Padre fumatore     | 59   | 130     | 189    |
| Padre non fumatore | 131  | 436     | 567    |
| Totale             | 190  | 566     | 756    |

Guardando questi dati, l'ipotesi di indipendenza tra sviluppo dell'asma e esposizione al fumo passivo durante la fase prenatale

Può essere rifiutata a un livello del 5%

12) I dati seguenti rappresentano i punteggi ottenuti da un campione di studenti a un esame di matematica

23, 22, 30, 21, 24 24, 25, 18, 18, 27, 19, 22, 28, 27, 20

Determinare il 35-esimo percentile

13) Dire di che tipo di variabili si tratta

Lunghezza della tibia: variabile quantitativa continua

Qualità dell'aria (ottima, molto buona, buona, discreta, accettabile, mediocre, scadente, inquinata, pessima):  
variabile qualitativa ordinale

Concentrazione di PM10 ( $\mu\text{g}/\text{m}^3$ ): variabile quantitativa continua

Temperatura atmosferica: variabile quantitativa continua

Numero di cuccioli in una cucciolata: variabile quantitativa discreta

gruppo sanguigno: variabile qualitativa categorica

14) E' noto che le mucche di razza Bruna Alpina producano meno latte delle mucche di razza Frisona. In un consorzio di allevamenti sono presenti in tutto 200 mucche di razza Bruna Alpina e 300 di razza Frisona. Per stimare la produzione media giornaliera di latte di una singola mucca si registra, per un anno, la produzione giornaliera di latte di un campione di 50 mucche. Come scegliereste questo campione?

Con un campionamento casuale a strati

15) I dati seguenti rappresentano i valori di trigliceridi nel sangue, espressi in mg/dl, di un campione di 12 adulti

174 196 220 183 194 200 192 200 200 199 213 193

Per questi valori la distanza interquartile

è uguale a 7,5

16) Nel seguente grafico ramo-foglia sono riportati i valori della pressione diastolica (mmHg) di un campione di 30 uomini

```

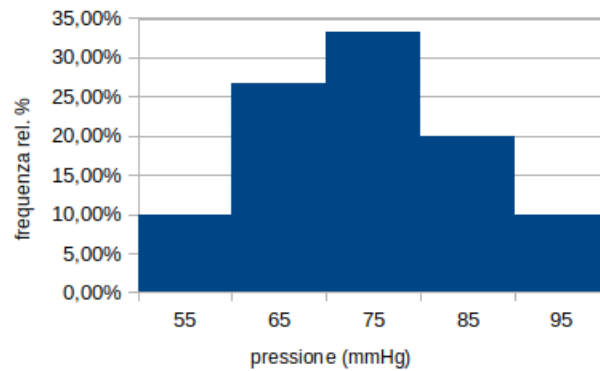
-----
5 4, 6, 8
-----
6 0,1, 2, 2, 3, 4, 5, 5
-----
7 0, 1, 2, 2, 4, 5, 5, 6,
  7, 8
-----

```

8 6, 7, 8, 9, 9, 9

9 3, 5, 8

Questi dati sono rappresentati nel seguente istogramma



17) Per questo insieme di dati

75 ,33, 55, 21, 46, 98, 103, 88, 35, 22, 29,73, 37, 101, 121, 144, 133, 52, 54, 63, 21, **7**

se sostituisco il valore **7** con il valore **20**

la mediana non cambia

18) I seguenti dati rappresentano il numero di cuccioli di topo partoriti da 16 topi femmina

7, 4, 8, 6, 10, 5, 11, 8, 15, 6, 9, 10, 11, 9, 13, 9

Dire quale delle seguenti affermazioni è corretta

Ordino i valori dal più piccolo al più grande. il primo quartile  $Q1$  è la media dei valori che occupano la quarta e la quinta posizione, quindi  $Q1 = (6+7)/2 = 6,5$ .  $Q3$  è la media dei valori che occupano la dodicesima e la tredicesima posizione, quindi  $Q3 = (10+11)/2 = 10,5$ . La distanza interquartile è  $Q3 - Q1 = 4$ .

La distanza tra 4 e  $Q1 = 6,5$  è 2,5 che è minore di  $1,5 \times (Q3 - Q1) = 6$ , quindi 4 non è un outlier.

La distanza tra 15 e  $Q3 = 10,5$  è 4,5 che è minore di  $1,5 \times (Q3 - Q1) = 6$ , quindi 15 non è un outlier. Non ci sono outlier.

19) La seguente tabella delle frequenze riporta i dati relativi alle età dei componenti di un'orchestra sinfonica giovanile

| <b>età</b> | <b>frequenza</b> |
|------------|------------------|
| 16         | 3                |
| 17         | 6                |
| 18         | 7                |
| 19         | 5                |
| 20         | 4                |

la media è circa uguale a 18.

la mediana è uguale a 18.