

## II ESERCITAZIONE

### ESERCIZIO 1

Quale percentuale di osservazioni sotto la curva normale standardizzata cade nell'intervallo compreso tra i valori  $z=-1$  e  $z=+1$ ?

- a) 66% circa
- b) 70,2% circa
- c) 68,2% circa
- d) 32% circa
- e) 31,6% circa

a      b      c      d      e

Soluzione 1: l'area prima di  $+1$  si trova con la formula  
 $=\text{DISTRIB.NORM}(1;0;1)$

e produce come risultato 0,841 pari allo 84,1% circa dell'area. L'area dopo 1 (che è uguale all'area prima di  $-1$ ) è  $1-0,841 = 0,159$ ; pertanto la soluzione è data da

$$0,841-0,159 = 0,682$$

pari al 68,2% circa.

Soluzione 2: usando le tavole si ha che l'area prima di 1 è 0.841 e il procedimento è analogo al precedente.

Soluzione 3: ricordiamo le regole di approssimazione.

tra  $\mu-\sigma$  e  $\mu+\sigma$  è compreso il 68% circa dei risultati;

tra  $\mu-2\sigma$  e  $\mu+2\sigma$  è compreso il 95% circa dei risultati;

tra  $\mu-3\sigma$  e  $\mu+3\sigma$  è compreso il 99,7% circa dei risultati.

In questo esercizio, avendo a che fare con la normale standard, abbiamo immediatamente che la risposta corretta è la c.

### ESERCIZIO 2

In un grande ospedale si vuole stimare il valore medio  $\mu$  della pressione sanguigna (mmHg) di tutti i pazienti dimessi dopo un intervento chirurgico negli ultimi due anni.

Si vuole estrarre un campione casuale di pazienti di dimensione 100 per calcolare  $s$ .

- i)  $s$  è un parametro relativo alla popolazione? **NO**
- ii) Qual è l'unità statistica? **Il paziente dimesso**  
 Qual è la variabile statistica? **La pressione**  
 Di che tipo di variabile si tratta? **Continua**

### ESERCIZIO 3

In uno studio si vuole confrontare la crescita media di piantine trattate con tre fertilizzanti, A, B e C.

Nella tabella seguente sono riportate le informazioni relative a un campione di sette piantine di pomodori di una data piantagione.

Tipo di fertilizzante	Altezza finale (cm)	Altezza iniziale (cm)
C	87	3
A	68	4
B	76	2
A	77	5
B	80	4
A	74	3
C	91	7

- i) Quali sono le popolazioni di riferimento?  
**Le popolazioni delle piantine trattate con i 3 fertilizzanti A, B e C.**
- ii) Quale è il parametro delle popolazioni che vogliamo stimare?  
**La crescita media  $\mu$**
- iii) Quale statistica viene usata per stimare tale parametro?  
**La media campionaria, ossia la statistica.**
- iv) Quale è l'unità statistica?  
**La piantina di pomodoro.**
- v) Si tratta di un esperimento o di uno studio di osservazione?  
**Esperimento.**

## ESERCIZIO 4

Il coefficiente di correlazione tra i punteggi a un test neurologico e a un test sull'ansia ha un valore alto e positivo, quale (o quali se ve ne è più di una) delle seguenti affermazioni è vera?

**a      b      c      d**

- a) quelli con un punteggio basso a un test tendono ad avere un punteggio basso nell'altro  
**Commento** Correlazione alta e positiva fa pensare ad una proporzionalità diretta crescente: se la  $x$  cresce, allora anche la  $y$  cresce; pertanto per valori bassi di un test ci aspettiamo valori bassi anche all'altro test.
- b) non si possono fare previsioni significative a partire da un test sull'altro test  
**Commento** Non possiamo fare previsioni perché non abbiamo i dati e perché non sappiamo se possiamo utilizzare un modello lineare.
- c) problemi neurologici causano ansia  
**Commento** Non abbiamo informazioni circa un rapporto di causa ed effetto.
- d) quelli con un punteggio basso a un test tendono ad avere un punteggio alto nell'altro  
**Commento** Contraddice la risposta a)

## ESERCIZIO 5

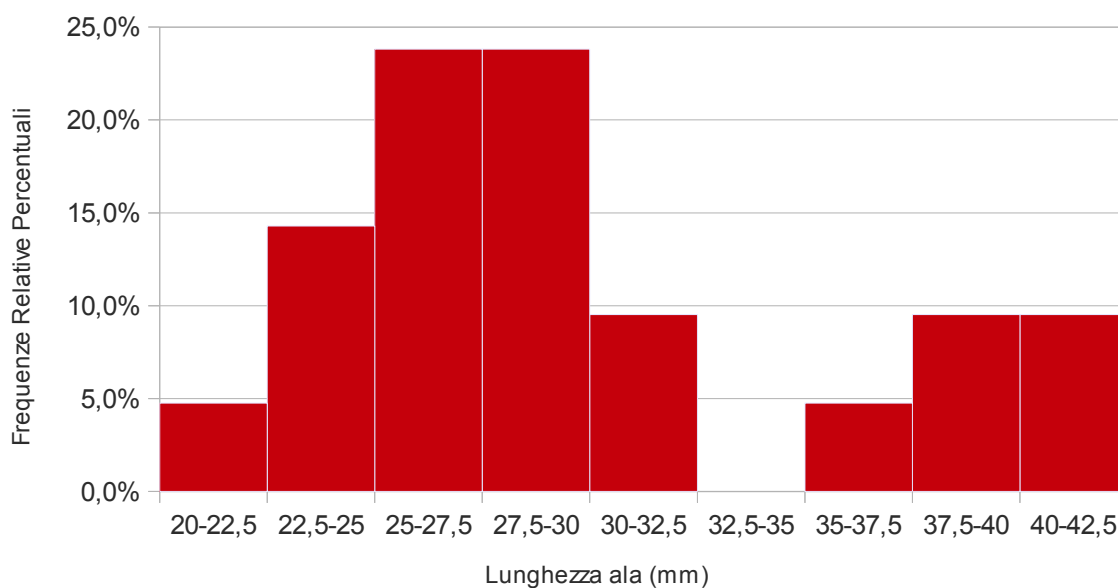
I dati che seguono riportano le misure in mm delle lunghezze di un'ala di 21 libellule.

41,1	42	23,5	24	40	30	30
30	32	32	24	25,2	25,3	25,3
25,4	26	35,4	29,2	39,2	28	21,5

- i) Per questi dati, considerate 9 classi di ampiezza pari a 2,5 a partire dal valore 20 e calcolate frequenze relative percentuali (con 1 cifra decimale) per ogni classe, riportandole qui sotto.

<b>Classi</b>	<b>Frequenze relative percentuali</b>
20-22,5	4,8%
22,5-25	14,3%
25-27,5	23,8%
27,5-30	23,8%
30-32,5	9,5%
32,5-35	0,0%
35-37,5	4,8%
37,5-40	9,5%
40-42,5	9,5%

- ii) Poi disegnate l'istogramma sulla base delle classi considerate, riportando il grafico sul foglio qui sotto. Specificate cosa riportate su entrambi gli assi.



- iii) Per questo campione di dati, calcolate il 10-mo percentile.

Per calcolare il p-esimo percentile si sceglie l'osservazione che occupa il posto dato dalla formula

$$p^*(n+1)=0,10*(21+1)=2,2$$

Il percentile cercato è dato, pertanto da un valore tra il secondo e il terzo valore, cioè tra 23,5 mm e 24 mm.

- iv) Verificate la normalità di queste misure usando gli indici descrittivi opportuni che conoscete.

Media = 29,96    Mediana = 29,2

Media e mediana sono molto vicine, abbiamo una distribuzione approssimativamente simmetrica.

Curtosi = -0,60

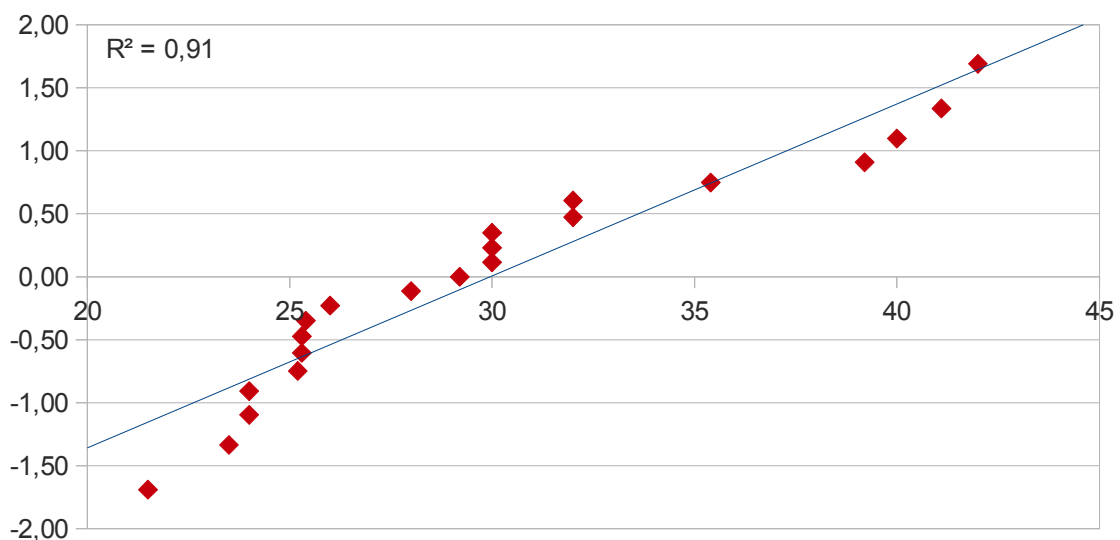
La curtosi negativa lascia pensare ad un comportamento iponormale, ossia una minore frequenza dei valori esterni.

Asimmetria = 0,75, si calcola con la funzione =ASIMMETRIA([Dati])

Siamo in presenza di una certa asimmetria, ad ogni modo, prima di concludere, disegniamo il normal plot.

- v) Verificate la normalità di queste misure costruendo un normal plot (che non dovrete riportare sul foglio). Questi dati si distribuiscono approssimativamente secondo una normale? Perché sì o perché no? Spiegate.

Ci sono scostamenti significativi dalla retta: le code si scostano dalla retta. Pertanto rifiutiamo l'ipotesi di normalità.



## ESERCIZIO 6

I dati che seguono riportano la massa dell'intestino ( $x$ ) e la massa del corpo ( $y$ ), misurate in grammi, di 10 cavie di 9 mesi.

Massa intestino ( $x$ )	Massa corpo ( $y$ )
28,0	130,0
30,0	140,0
30,4	140,7
31,0	170,0
31,0	149,0
36,0	140,0
40,0	140,0
38,0	156,0
31,5	131,0
30,1	133,5

Si vuole studiare se esiste una relazione tra la massa del corpo e quella dell'intestino.

- i) Calcolate il coefficiente di correlazione e valutate la forza della relazione tra le due variabili.

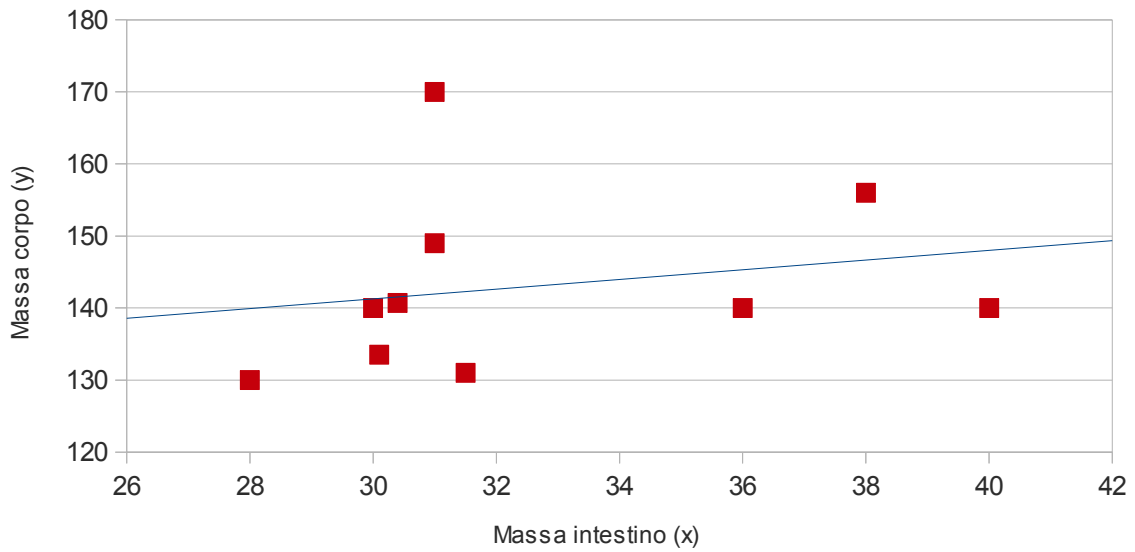
$R=0.22$ ; relazione molto debole

- ii) Quale percentuale di variabilità delle  $y$  è spiegata dal modello di regressione?

Il coefficiente di determinazione è pari a 4,64%

- iii) C'è un'osservazione che si discosta molto dalle altre. Spiegate se si tratta di un outlier o di un'osservazione influente

Guardando lo scatterplot notiamo come c'è una osservazione che si discosta dalle altre:



Si discosta nella direzione delle y e pertanto (31;170) è un outlier.

iv) Calcolate il residuo corrispondente a tale osservazione

28,05

v) Si può concludere che il modello di regressione è un buon modello per questi dati? Spiegate.

No, non è un buon modello in quanto abbiamo un coefficiente di determinazione molto basso.