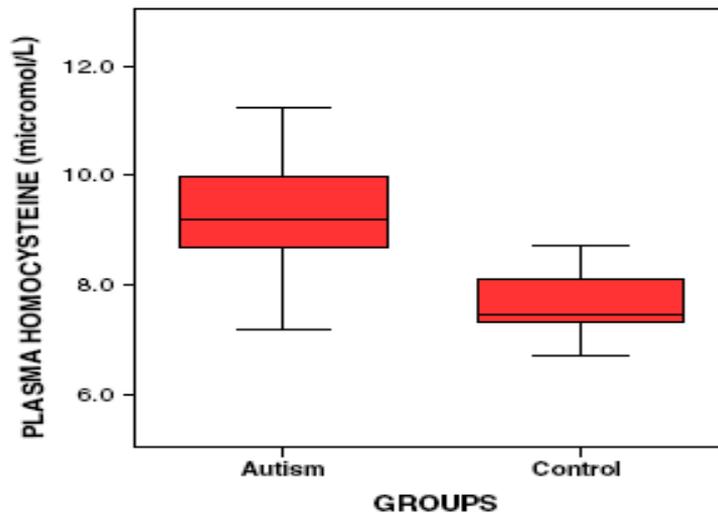


ALCUNI ESERCIZI ASSEGNATI (3 e 6 crediti) NELLA SESSIONE DI LUGLIO CON RELATIVE SOLUZIONI
ATTENZIONE I VALORI OTTENUTI NEI CALCOLI (OTTENUTI CON UN SOFTWARE STATISTICO) POSSONO DIFFERIRE DI POCO DA QUELLI OTTENUTI CON OPEN OFFICE

ESERCIZIO

In uno studio alcuni ricercatori hanno confrontato il numero di omocisteine nel plasma (micromol/L) in bambini affetti da autismo e in un gruppo di bambini sani (Controllo). Nei box-plots che seguono sono riportate le distribuzioni del numero di omocisteine nel plasma nei bambini affetti da autismo e nei sani.



i) Spiegate se si tratta di uno studio di osservazione o di un esperimento.

Studio di osservazione, perché non viene imposto nessun trattamento alle unità campionarie

ii) Confrontate le due distribuzioni di dati spiegando tutte le differenze che riscontrate in termini di forma, di misure di posizione e variabilità.

1° quartile autistici > 3° quartile controllo
 autistici: distr. asimmetrica con scarto interquartile minore di 2
 controllo: distr. asimmetrica con scarto interquartile circa pari ad 1

ESERCIZIO

Se il valore del coefficiente di determinazione R^2 è positivo se ne deduce una relazione lineare positiva tra le due variabili.

V **F(x)**

La varianza è sempre maggiore della media.

V **F(x)**

Se un qualunque insieme di dati è standardizzato, la media dei valori standardizzati vale 0.

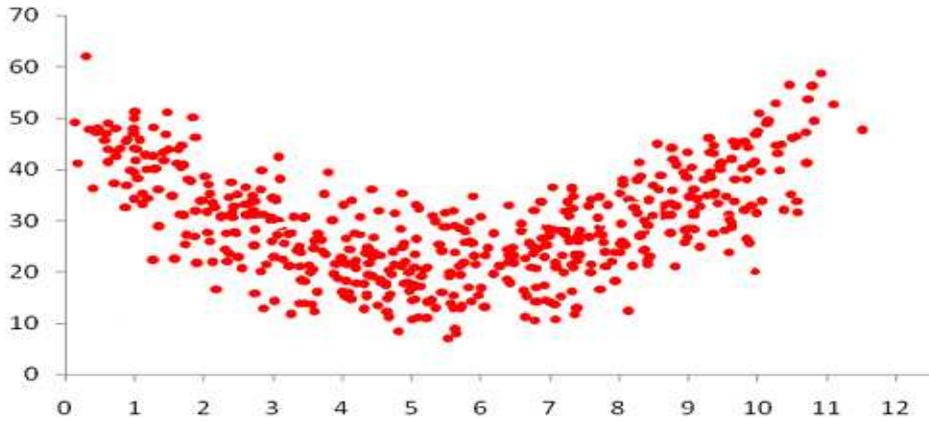
V(x) **F**

Se si cambia il segno a ognuno dei dati di una serie, la deviazione standard cambia segno.

V **F(x)**

ESERCIZIO

Spiegate se vi sembra che si possa dire che il diagramma di dispersione che segue lasci presagire una dipendenza lineare prima inversa e poi diretta tra le due variabili in studio.



NO: il diagramma di dispersione mostra una relazione curvilinea tra le 2 variabili oggetto di studio

ESERCIZIO

In un istogramma la frequenza relativa in ogni un rettangolo del grafico si ottiene facendo il rapporto tra l'altezza della barra e l'ampiezza della base. $V = \frac{F(x)}{F}$

ESERCIZIO

Nella seguente distribuzione di frequenze, il valore nella casella mancante è :

- a) 34
- b) 40
- c) 32(x)
- d) nessuno dei precedenti

Modalità	Frequenze assolute	Frequenze percentuali
A	13	10,8%
B	27	22,5%
C		26,7%
D	48	40%

ESERCIZIO

Nella tabella che segue sono riportati alcuni dati biometrici di 17 individui.

Genere	Altezza cm	Num scarpe	Colore occhi
M	185	39	marrone
F	165	39	marrone
F	165	40	marrone
F	170	36	marrone
F	175	42	verde
M	176	42	marrone
F	171	38	marrone
M	170	41	marrone
F	163	39	verde
F	170	39	marrone
F	160	36	marrone
F	174	41	blu
F	164	36	blu
M	177	41	marrone
M	180	44	verde
F	161	36	marrone
M	162	41	marrone

i) Disegnate i grafici appropriati per ciascuna delle seguenti variabili: altezza, numero di scarpa, e colore degli occhi (per l'istogramma dell'altezza considerate classi di ampiezza pari a 4 cm a partire dal valore 158cm) e riportate i 3 grafici sull'ultimo foglio.

ISTOGRAMMA DELL' ALTEZZA

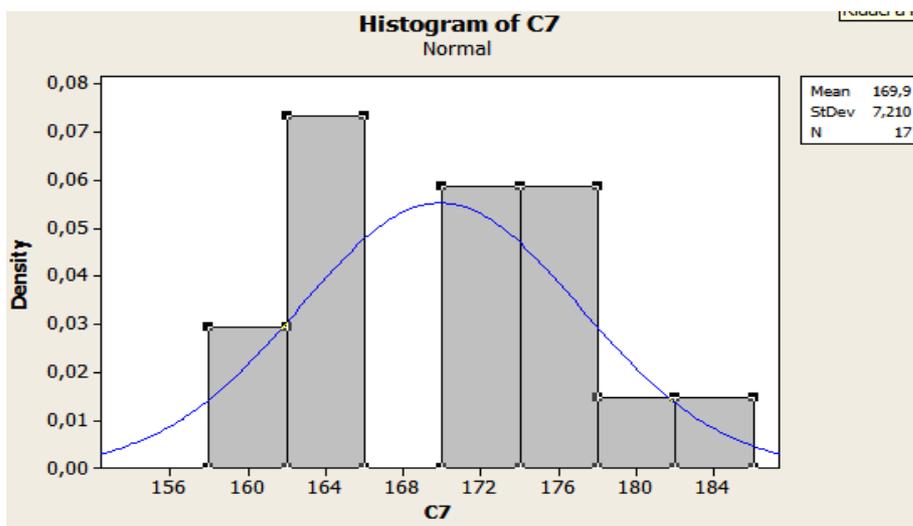
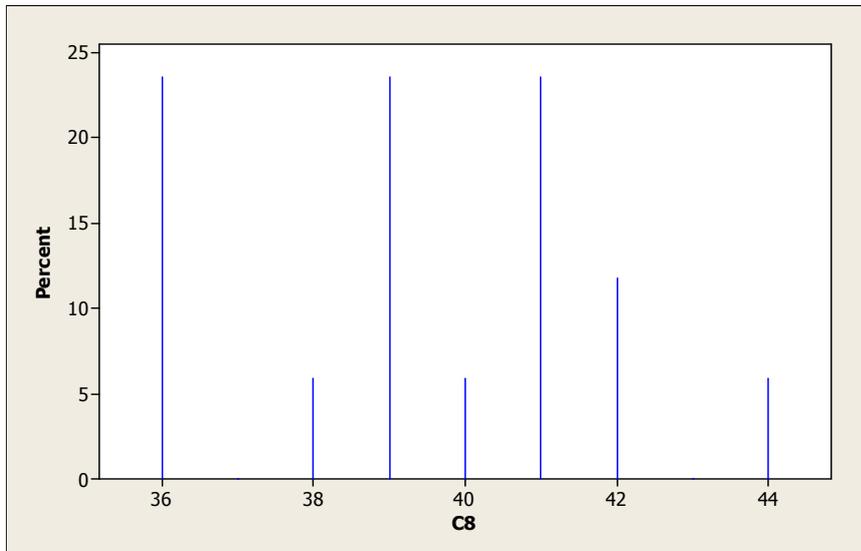


DIAGRAMMA A SEGMENTI DEL N° DI SCARPA



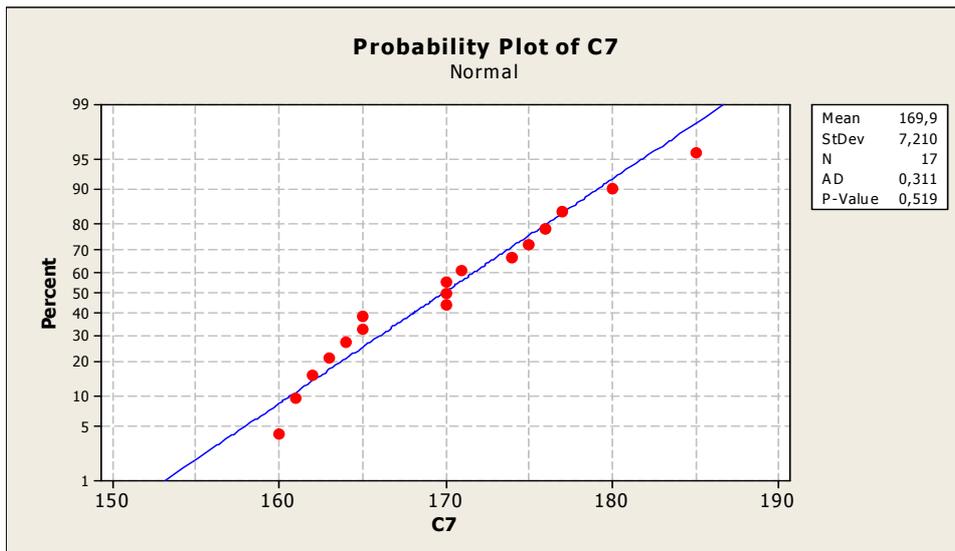
PER IL COLORE DEGLI OCCHI SI DISEGNA UN GRAFICO A TORTA

ii) Per i grafici dell'altezza e del numero di scarpa commentate la forma della distribuzione. Per verificare eventuali asimmetrie potete usare tutti gli indici che ritenete più opportuni. Riportate commenti e risultati.

1) altezza: Approssimativamente normale. Media = 169,8 Mediana = 170
Coeff. asimmetria = 0,44

2) n° scarpa: Multimodale

iii) Disegnate il normal plot per la variabile altezza senza riportarlo sul foglio. Osservando il grafico si può dire che tale variabile segue approssimativamente una distribuzione normale? Spiegate



Il normal plot è stato disegnato con un software statistico. Si può vedere che la variabile altezza segue una distribuzione approssimativamente normale. Si giunge alla stessa conclusione anche usando la procedura che avete imparato usando Open Office.

iv) Calcolate per ciascuna delle 3 variabili considerate le misure di posizione e di dispersione opportune.

altezza: $m=169,9$ mediana= 170 dev. st. = $7,21$ coeff. asim. = $0,44$ curtosi = $-0,56$

n° scarpe: $m=39,4$ dev. st. = $2,4$ coeff. asim. = $-0,07$ curtosi= $-0,75$

colore occhi: moda (classe modale) = marrone

v) Dopo aver calcolato la tabella delle frequenze relative percentuali (oppure delle frequenze percentuali cumulate) per la variabile altezza, dite quale percentuale di valori della distribuzione è, approssimativamente, compresa fra $\bar{x} - s$ e $\bar{x} + s$. Si ricorda che \bar{x} e s sono la media e la deviazione standard del campione.

70,57%

vi) In una distribuzione normale quale percentuale di valori è compresa tra $\mu - \sigma$ e $\mu + \sigma$?

68,2%

ESERCIZIO

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)
Tipo C	87	3
Tipo A	68	4
Tipo B	76	2
Tipo A	77	5
Tipo B	80	4
Tipo A	74	3
Tipo C	91	7

i) Qual è la popolazione di riferimento?

tutte le piantine della piantagione

ii) Qual è l'unità statistica?

piantina

iii) Descrivete tutte le variabili ed i rispettivi valori o modalità assunti nel campione.

Fertilizzante: A, B, C

Altezza finale: valori osservati 68 74 76 77 80 87 91

Altezza iniziale: valori osservati 2 3 4 5 7

iv) Si vuole vedere se c'è una relazione tra l'altezza iniziale e l'altezza finale delle piantine. Stabilite, con gli strumenti che conoscete, se tale relazione esiste.

$r = 0,482$ retta di regressione: $h \text{ finale} = 69,8 + 2,31 h \text{ iniziale}$

v) Stabilite se un modello di regressione descrive bene i dati.

$R^2 = 23,3\%$ \rightarrow Un modello di regressione non descrive bene i dati

vi) C'è un valore y che si discosta dagli altri e che ha un residuo grande. Di quale valore si tratta? Qual è il suo residuo?

$y=4,68$ residuo=-11

vii) E' possibile eliminare l'osservazione corrispondente a tale valore per migliorare l'adattamento del modello di regressione ai dati? Spiegate.

NO Non si può eliminare un'osservazione campionaria se non si hanno motivi fondati