

I ESERCITAZIONE

ESERCIZIO 1

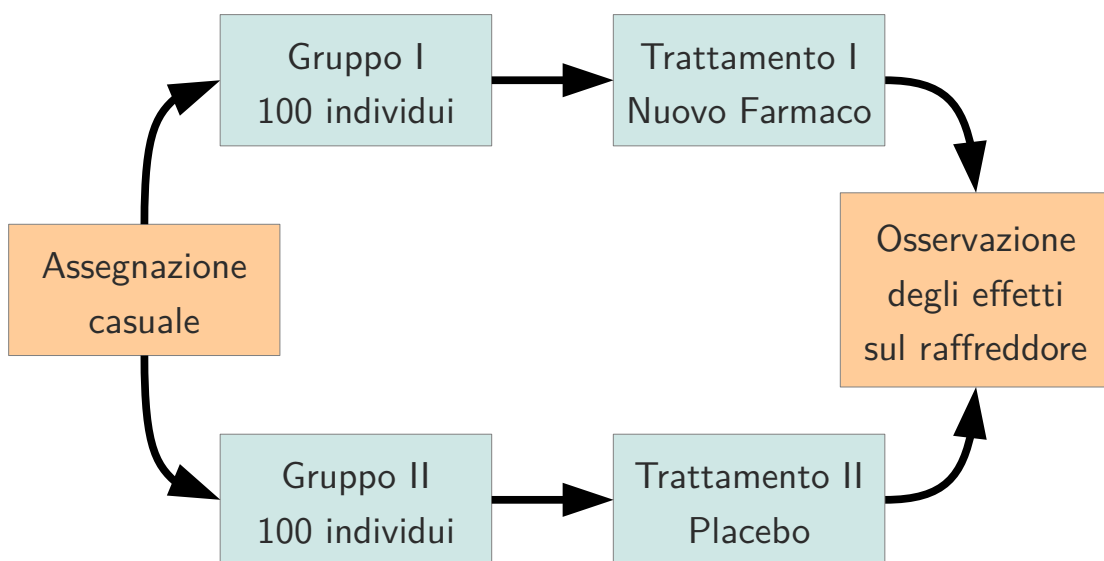
Si vuole testare un nuovo farmaco contro il raffreddore.

Allo studio partecipano 200 soggetti sani della stessa età e dello stesso sesso e con caratteristiche simili.

- i) Che tipo di disegno degli esperimenti usereste? Spiegate, con precisione, i passi necessari per svolgere l'esperimento. Se volete potete disegnare un grafico.

Usiamo un esperimento completamente randomizzato.

Si considerano 2 trattamenti: farmaco e placebo. I soggetti devono essere assegnati a caso ai trattamenti.



Numerare i soggetti con numeri da 001 a 200.

I primi 100 numeri generati dalle tabelle dei numeri casuali (considerando blocchi di 3 cifre) individuano i 100 soggetti da inserire nel Gruppo I; i restanti 100 vanno inseriti nel Gruppo II.

Esempio: dalla tavola seguente, raggruppiamo i numeri a gruppi di 3 cifre e scartiamo quelli maggiori di 200.

36	08421	44753	77377	28744	75592	08563	79140	92454
37	53645	66812	61421	47836	12609	15373	98481	14592
38	66831	68908	40772	21558	47781	33586	79177	06928
39	55588	99404	70708	41098	43563	56934	48394	51719
40	12975	13258	13048	45144	72321	81940	00360	02428
41	96767	35964	23822	96012	94591	65194	50842	53372
42	72829	50232	97892	63408	77919	44575	24870	04178
43	88565	42628	17797	49376	61762	16953	88604	12724
44	62964	88145	83083	69453	46109	59505	69680	00900
45	19687	12633	57857	95806	09931	02150	43163	58636

Inseriamo nel Gruppo I i soggetti:

084 126 142 147 091 ...

ii) Perché è necessario che i soggetti siano il più simile possibile tra loro?

Per evitare l'effetto di variabili nascoste confondenti.

ESERCIZIO 2

Nella tabella che segue sono riportate le dosi (in mg) di un farmaco antipertensivo e le relative diminuzioni di pressione sistolica misurate in 8 pazienti dopo averlo ingerito.

Dose (x)	Diminuzione (y)
8	12
20	24
30	6
10	22
18	21
14	14
20	25
16	26

i) Calcolate a quale dose corrisponde il residuo positivo più piccolo nel modello di regressione lineare.

10

- ii) Riportate il valore di tale residuo **1,46**
- iii) Riportate l'interpretazione del coefficiente di determinazione

$$r^2 = \frac{\text{Variabilità spiegata}}{\text{Variabilità totale}} = \text{coefficiente di determinazione}$$

- iv) Sul grafico di dispersione appare un'osservazione lontana dalle altre e con il residuo più alto

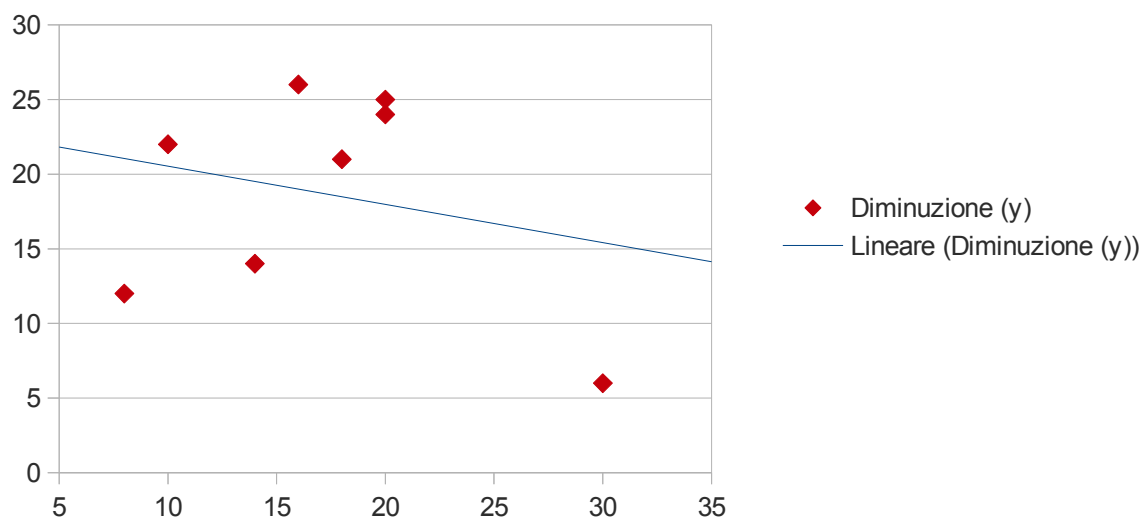
1) A quali valori corrisponde tale osservazione? **(30;6)**

2) Si tratta di un outlier o di un'osservazione influente?

È un'osservazione influente. Infatti da una parte si ha che il comportamento anomalo si verifica per le ascisse; d'altra parte, eliminando questa osservazione il risultato cambia profondamente. Diamo uno sguardo ai grafici:

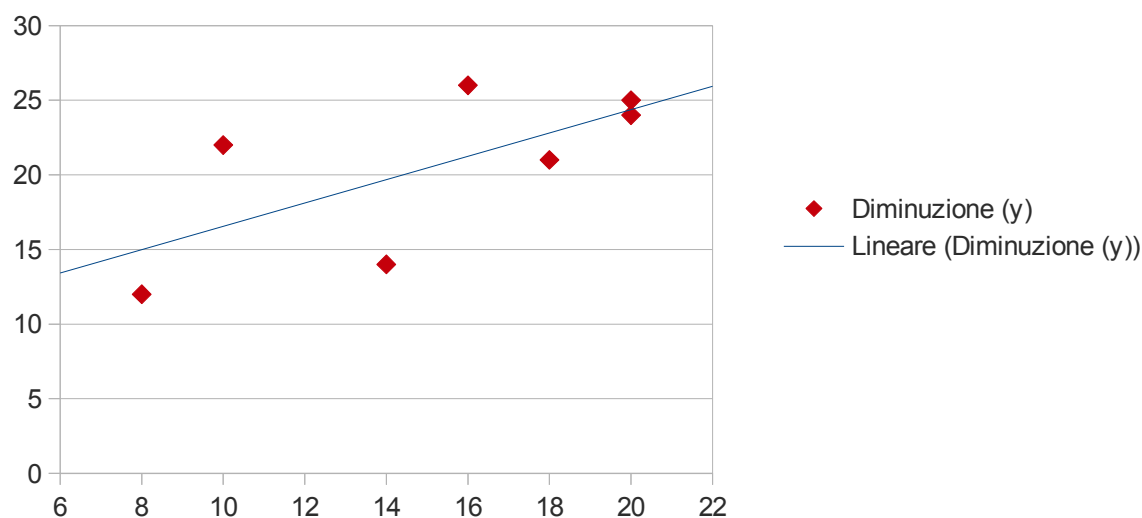
Con l'osservazione (30;6)

Diagramma di dispersione



Senza l'osservazione (30;6)

Diagramma di dispersione



- v) Spiegate se si può dire che questa retta di regressione rappresenta un buon modello per i dati.

No, il coefficiente di determinazione è 5,9%

La regressione lineare non è un buon modello

- vi) Possiamo dire che la dose di farmaco in questione può essere messa in relazione con la corrispondente diminuzione di pressione arteriosa? **V** **F**

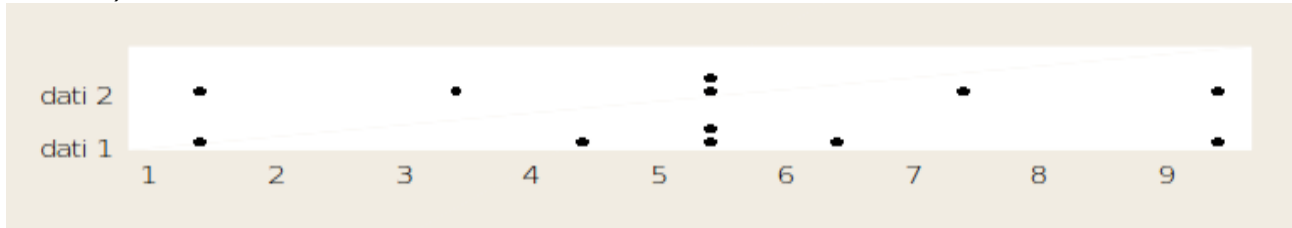
- vii) Disegnate il grafico dei residui della regressione (senza riportarlo sul compito) e dite quale variabile riportate sull'asse delle ascisse e quale sull'asse delle ordinate.

Asse Y: "Residui"

Asse X: "Valori osservati (x)" oppure "Valori stimati (y)"

ESERCIZIO 3

Nei due diagrammi che appaiono sotto, ogni punto rappresenta un'osservazione. Osservate i diagrammi per rispondere alle domande (non dovete fare nessun calcolo).



i) I “dati 1” hanno la media più piccola.

V F

Data la simmetria centrale, la media pare essere la stessa

ii) Il terzo quartile dei “dati 1” è maggiore del terzo quartile dei “dati 2”

V F

Dovrebbe essere più piccolo

iii) I “dati 1” hanno la deviazione standard più grande

V F

No, è più piccola: i “dati 1” si scostano meno dalla media

ESERCIZIO 4

Qual percentuale del campione di osservazioni sotto la curva normale standardizzata cade nelle 2 code fuori dai valori $z=-2,5$ e $z=2,5$?

- a) 0,62% circa
- b) 98,76% circa
- c) 1,24% circa
- d) 49,38% circa
- e) 50,62% circa

a b c d e

Soluzione 1: l'area prima di $-2,5$ si trova con la formula

$$= \text{DISTRIB.NORM}(-2,5;0;1)$$

e produce come risultato 0,0062 pari allo 0,62% circa dell'area. Per simmetria l'area dopo 2,5 è uguale; pertanto la soluzione è data da

$$0,62\% + 0,62\% = 1,24\%$$

Soluzione 2: usando le tavole si ha che l'area prima di 2,5 è 0.99379

Quella dopo 2.5 è

$$1-0.99379 = 0,00621$$

pari allo 0,62% circa. Pertanto la soluzione è data da

$$0,62\%+0,62\% = 1,24\%$$

ESERCIZIO 5

I dati che seguono riportano le misure in mm delle lunghezze della zampa di 21 tordi.

22,0	22,0	23,5	24,0	21,1	30,0	30,0
30,0	32,0	32,0	24,0	25,2	25,3	25,3
25,4	26,0	35,4	29,2	29,1	28,0	20,6

- i) Per questi dati, considerate 9 classi di ampiezza pari a 2 a partire dal valore 19 e calcolate le densità (con due cifre decimali) per ogni classe, riportandole qui sotto.

Classi	Densità
19-21	0,02
21-23	0,07
23-25	0,07
25-27	0,12
27-29	0,02
29-31	0,12
31-33	0,05
33-35	0,00
35-37	0,02

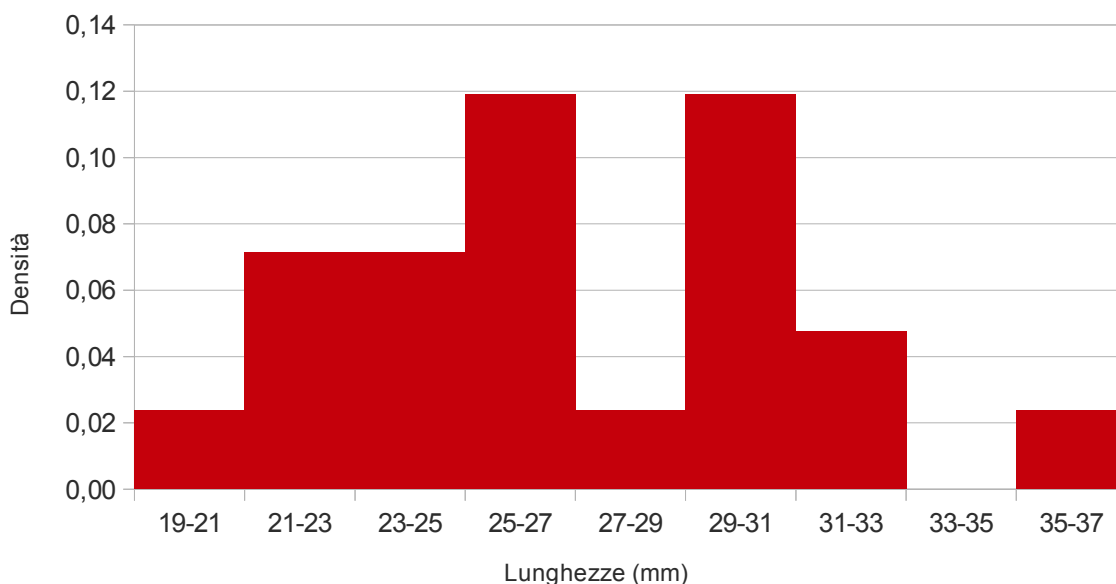
Ricordiamo che le densità si trovano calcolando:

Frequenza

Frequenza relativa = [Frequenza]/[Numero di elementi del campione]

Densità = [Frequenza relativa]/[Ampiezza classe]

- ii) Poi disegnate l'istogramma, riportando il grafico sul foglio qui sotto. Scegliete una scala tale che il grafico sia leggibile.



- iii) Dite quali sono tutti gli altri grafici che si possono usare per descrivere questa distribuzione di dati, senza disegnarli.

Ramo-Foglia, Box Plot

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

Media = 26,67,

Deviazione Standard = 4,02

- v) Supponete che questo campione di dati provenga da una distribuzione approssimativamente normale con media e deviazione standard pari a quelle calcolate sul campione (vedi punto precedente).

- 1) Quale è la percentuale di tordi con lunghezza della zampa maggiore di 21 mm?

0,921

Infatti, la percentuale di tordi con lunghezza della zampa inferiore a 21mm è data dalla formula

=DISTRIB.NORM(21;[Media];[Deviazione standard])

che ha come risultato 0,079

La soluzione è quindi data da $1-0,079 = 0,921$

2) Quale valore della lunghezza della zampa corrisponde al 60esimo percentile?

27,69

Infatti, la formula che calcola il risultato è

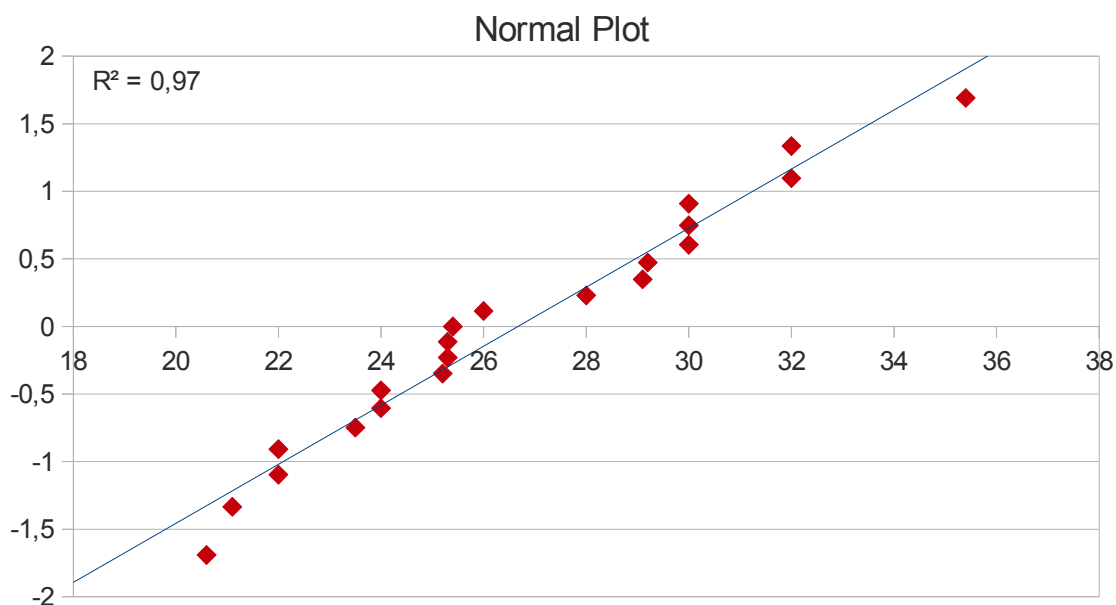
$=\text{INV.NORM}(0,6;[\text{Media}];[\text{Deviazione standard}])$

che produce 27,69

vi) 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.

I dati si dispongono nelle vicinanze della retta con un coefficiente di determinazione pari a 0,97.

Un'occhiata al normal plot:



Una curtosi pari a -0,58, calcolata con la formula $=\text{CURTOSI}([\text{Dati}])$, e una asimmetria di 0,36, calcolata con la formula $=\text{ASIMMETRIA}([\text{Dati}])$, non creano problemi e accettiamo pertanto l'ipotesi di normalità.