**DATI:**

A <- read.table ("clipboard")

1.67 61 04 0

1.78 62 26 0

1.78 72 05 0

1.84 75 2 0

1.82 76 12 0

1.80 72 2 0

1.95 91 3 0

1.81 80 9 0

1.80 70 9 0

1.86 77 6 0

1.68 65 9 0

1.84 64 8 0

1.72 62 1 0

1.76 65 26 0

1.82 74 4 0

1.86 86 3 0

1.84 75 14 0

1.73 65 19 0

1.72 57 25 1

1.62 56 2 **1**

1.70 59 3 1

1.61 46 15 1

1.70 64 22 1

1.60 63 26 1

1.61 60 16 1

1.60 55 28 1

1.65 50 12 1

Alt = A[,1]

Peso = A[,2]

Data = A[,3]

Sesso = A[,4]

plot(Alt,Peso)

abline(lm(Peso~Alt),col="red")



Esercizio 1. Colorare i maschi di blu, le femmine di rosso.

plot(Alt,Peso,col= "blue")

lines(Alt[19:27],Peso[19:27],col= "red",type="p")

abline(lm(Peso~Alt),col="black")



Esercizio 2. Oltre alla retta nera, disegnare anche la retta rossa delle sole femmine e quella blu dei soli maschi.

abline(lm(Peso[1:18]~Alt[1:18]),col="blue")

abline(lm(Peso[19:27]~Alt[19:27]),col="red")



1. **CORRELAZIONE**

cor(Alt,Peso)

Il risultato è

> cor(Alt,Peso)

[1] 0.8630517

È elevato? Quanto viene il coefficiente di correlazione di stringhe indipendenti della stessa lunghezza?

Vediamo come si può esaminare statisticamente la correlazione tra campioni indipendenti , per capire cosa possiamo aspettarci. Il numero n (fatelo variare) è la numerosità del campione, da cui dipenderà fortemente il grafico; il numero N invece è ausiliario, è solo la precisione con cui cerchiamo il grafico.

n=27; N=100000

COR = 1:N

for (i in 1:N) {

x= rnorm(n)

y= rnorm(n)

COR[i] = cor(x,y)

 }

hist(COR,100)



Osserviamo che il valore della tabella

> cor(Alt,Peso)

[1] 0.8630517

È completamente esterno a questo grafico e quindi mostra una sicura correlazione.

Vediamo come cambia il grafico con la numerosità del campione sperimentale:

n=94; N=10000

COR = 1:N

for (i in 1:N) {

x= rnorm(n)

y= rnorm(n)

COR[i] = cor(x,y)

 }

hist(COR,50)



Nota: si tratta di un modo di eseguire test statistici, senza il contorno della teoria: si confronta una distribuzione di riferimento, legata all’ipotesi da mettere in discussione (la cosiddetta ipotesi nulla) con un valore sperimentale; se il valore sperimentale è anomalo rispetto alla distribuzione di riferimento, si rifiuta l’ipotesi nulla.

Il p-value del valore sperimentale rispetto alla distribuzione di riferimento è l’area sotto tale curva a partire dal valore sperimentale.

1. **CORRELAZIONE**

I seguenti esempi illustrano l’interpretazione geometrica del coefficiente di correlazione.

X = 1:10

Y= 5\*X

cor(X,Y)

plot(X,Y)

-------

X = 1:10

Y= - 7\*X+2

cor(X,Y)

plot(X,Y, type="l")

--------

X = 1:20

Y= 5\*X + 10\*rnorm(20)

cor(X,Y)

plot(X,Y)

-------------

X=rnorm(100)

Y=rnorm(100)

cor(X,Y)

plot(X,Y)

abline(lm(Y~X),col="black")

1. **QUANTILI, CUMULATIVA EMPIRICA**

X=rnorm(100)

ecdf(X)

plot(ecdf(X))

X=rnorm(10000)

plot(ecdf(X))

X1= seq(-5,5,0.01)

Y= pnorm(X1)

lines(X1,Y, col="red")

Oppure:

XX=sort(X)

Y= pnorm(XX)

lines(XX,Y, col="red")

