1. ***RICHIAMO: CARICAMENTO DI TABELLE***

Se abbiamo una tabella, scritta su un foglio, la procedura è:

1. Scrivere (o copiare) su R il comando A <- read.table ("clipboard") senza dare d’invio.
2. Andare sul foglio dove c’è la tabella, selezionarla e fare “copia”
3. Tornare col cursore su R e dare “invio”.
4. **UN ESEMPIO REALE**

Con la procedura appena vista, caricare su R la seguente tabella:

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

Isoliamo con nomi facili le diverse colonne (in altri casi i nomi ci sono già nella tabella di partenza, qui no).

Alt = A[,1]

Peso = A[,2]

Data = A[,3]

Sesso = A[,4]

cor(Alt,Peso)

plot(Alt,Peso)

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



Quindi approssimativamente c’è una regola che dice il peso data l’altezza, con una banda di confidenza.

**PROBLEMA**: Come si può determinare e disegnare una banda di confidenza?

Potremmo immaginare che la regola dipenda dal sesso. Vediamo graficamente i due gruppi colorati diversamente (18 maschi, 9 femmine):

plot(Alt[1:18],Peso[1:18],col="blue")

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

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

Non ha funzionato. Il comando lines non è proprio esattamente come plot: di default produce linee invece che punti.

plot(Alt[1:18],Peso[1:18],col="blue")

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

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

Di nuovo non ha funzionato. Il comando lines non allarga la finestra. Bisogna quindi settare la finestra all’origine.

plot(c(1.50,2),c(50,100),type="n")

plot(Alt,Peso)

lines (Alt[1:18],Peso[1:18], type="p",col="blue")

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

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



Non c’è alcuna evidenza grafica che il rapporto peso/altezza cambi tra maschi e femmine.

Infine, esaminiamo il legame tra altezza e data di nascita.

cor(Alt,Data)

plot(Alt,Data)

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

cor(Peso,Data)

plot(Peso,Data)

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

cor(Peso[1:18],Data[1:18])

plot(Peso[1:18],Data[1:18])

abline(lm(Data[1:18]~Peso[1:18]),col="red")

cor(Peso[19:27],Data[19:27])

plot(Peso[19:27],Data[19:27])

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

I risultati globali e dei maschi sono sorprendenti. Analizziamoli allora anche col metodo visto nell’esercitazione precedente.

n=27; N=50000

COR = 1:N

for (i in 1:N) {

x= rnorm(n)

y= rnorm(n)

COR[i] = cor(x,y)

 }

hist(COR,50)

1. **ESERCIZI**
2. Come si può determinare e disegnare una banda di confidenza, nella regressione?
3. Come dare una risposta scientifica (numerica) e non solo graficamente intuitiva, riguardante la bontà della correlazione 0.863 e l’anomalia della correlazione -0.41?