1. **SULLA TABELLA GIA’ ESAMINATA, con la variante header=TRUE**

Spesso è desiderabile che le tabelle abbiano i nomi nelle colonne. I diversi comandi sono:

A <- read.table ("clipboard")

A <- read.table ("clipboard", header=TRUE)

A <- read.table ("clipboard", header=T)

Prima di tutto, se una tabella senza nomi è già stata caricata, ad esempio

B<- read.table ("clipboard")

1.67 61 04 0

1.78 62 26 0

1.78 72 5 0

1.84 75 2 0

1.82 76 12 0

Allora basta scrivere

colnames(B)=c("Alt","Peso","Data","Sesso") % attribuisce i nomi alle colonne di una tabella già caricata

Invece, se si possiede su un foglio di testo una tabella coi nomi, del tipo

Alt Peso Data Sesso

1.67 61 04 0

1.78 62 26 0

1.78 72 5 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

basta caricarla con

A <- read.table ("clipboard", header=TRUE)

1. **REGRESSIONE COI NOMI DELLA TABELLA**

Ricordiamo che in precedenza, per eseguire una regressione, avevamo prima dato i nomi alle colonne. Ora ce l’hanno già. Allora basta scrivere

RRR = lm(Peso~Alt, data=A)

Chiedendo summary(REG) si possono vedere i risultati. Questo metodo però ha il difetto che se si vuole un plot delle prime due colonne (ad es. per aggiungere la retta di regressione), il software non riconosce comandi del tipo plot(Alt, Peso). Bisogna scrivere

plot(A$Alt, A$Peso)

abline(RRR)

In generale, cioè, le colonne della tabella si chiamano A$...

1. **STANDARDIZZARE UNA TABELLA**

Questa operazione non è indispensabile. Ogni tanto torna utile, quindi vediamola.

Innanzi tutto, dato un vettore, ad esempio

X=rnorm(100,47,10)

lo si può standardizzare, coi comandi

X.st = (X-mean(X))/sd(X)

Se ora chiediamo media e deviazione del nuovo vettore, troviamo 0 ed 1:

> mean(X.st)

[1] -1.884842e-16

> sd(X.st)

[1] 1

> round(mean(X.st))

[1] 0

Si osservi il funzionamento del comando round(,k), in alcuni esempi.

Sia ora A una tabella. La sua standardizzata è intesa per colonne: si standardizza ciascuna colonna. Per standardizzarla, basta usare i comandi:

A.st=A

A.st[,1] = (A[,1]-mean(A[,1]))/sd(A[,1])

A.st[,2] = (A[,2]-mean(A[,2]))/sd(A[,2])

………

**Esercizio**: scrivere un ciclo di for. Oppure altri metodi.

E così via per le altre colonne. Controlliamo:

> A.st

Alt Peso Data Sesso

1 -0.8049134 61 4 0

2 0.3444100 62 26 0

3 0.3444100 72 5 0

4 0.9713137 75 2 0

5 0.7623458 76 12 0

6 0.5533779 72 2 0

7 2.1206371 91 3 0

8 0.6578619 80 9 0

9 0.5533779 70 9 0

10 1.1802816 77 6 0

11 -0.7004294 65 9 0

Ecc.

Oppure:

A.st[,1] = round((A[,1]-mean(A[,1]))/sd(A[,1]),2)

> A.st

Alt Peso Data Sesso

1 -0.80 61 4 0

2 0.34 62 26 0

3 0.34 72 5 0

4 0.97 75 2 0

5 0.76 76 12 0

6 0.55 72 2 0

7 2.12 91 3 0

8 0.66 80 9 0

9 0.55 70 9 0

10 1.18 77 6 0

11 -0.70 65 9 0

Che è molto più leggibile.

1. **ESAMI PRELIMINARI DELLA TABELLA “INDICATORI DI BENESSERE”**

La seguente tabella è ripresa più volta anche nelle dispense teoriche, per esemplificare alcune nozioni. Riguarda un misto di indicatori relativi al sistema sanitario, alla salute, alle condizioni economiche delle persone o famiglie. I dati sono già standardizzati e sono stati troncati alla terza cifra decimale.

IB <- read.table ("clipboard", header=TRUE)

PLIC SC SA.SC TD TMI

Piem 0.088 0.471 -0.707 -0.607 -0.3950

Vaos -1.545 0.348 -0.642 -0.813 1.5780

Lomb 0.202 1.397 -0.836 -0.790 -0.5380

TrAA 0.677 0.435 -1.269 -0.966 -0.0750

Vene 0.088 1.334 -1.210 -0.848 -0.4970

FrVG 0.639 -0.005 -1.028 -0.804 -1.3010

Ligu 1.190 -0.247 0.470 -0.429 -0.3540

EmRo 0.658 1.177 -1.315 -0.863 -0.3470

Tosc 0.126 1.092 -0.795 -0.644 -1.3550

Umbr -1.431 0.675 -0.140 -0.524 -1.2870

Marc 0.278 1.090 -0.265 -0.702 -0.0006

Lazi 2.329 0.546 -0.080 -0.113 -0.0140

Abru 0.335 -0.373 0.402 -0.456 0.0400

Moli 0.658 -1.289 0.065 0.451 -1.1510

Camp -1.811 -1.314 2.031 1.664 0.4140

Pugl -0.766 -0.926 1.038 0.648 1.1090

Basi -0.747 -1.154 0.661 0.844 2.0010

Cala -0.500 -1.727 1.571 2.153 0.6320

Sici -0.918 -1.130 1.332 1.517 1.7830

Sard 0.449 -0.403 0.717 1.285 -0.2380

Dalla sola lettura della tabella è difficile estrarre informazioni. Con i seguenti primi comandi possiamo farci una prima idea dei legami tra le variabili:

> round(cor(IB),2)

PLIC SC SA.SC TD TMI

PLIC 1.00 0.32 -0.41 -0.37 -0.44

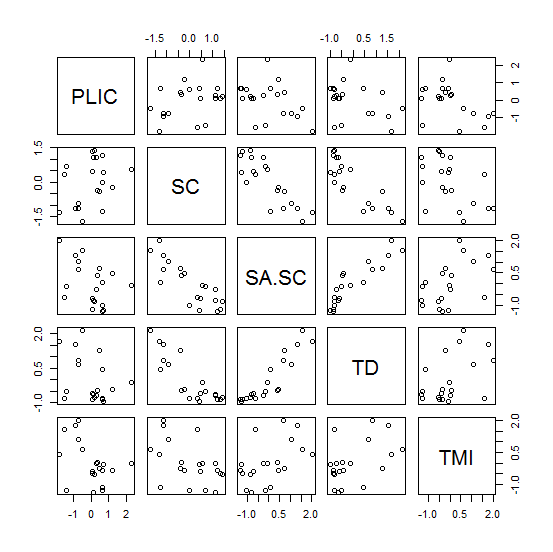
SC 0.32 1.00 -0.84 -0.85 -0.48

SA.SC -0.41 -0.84 1.00 0.91 0.51

TD -0.37 -0.85 0.91 1.00 0.49

TMI -0.44 -0.48 0.51 0.49 1.00

plot(IB)



Vediamo più da vicino una coppia di variabili molto collegate:

plot(IB$TD, IB$SA.SC)

abline(lm(SA.SC~TD,data=IB))

