1. **ISTOGRAMMI E GRAFICI DI ALCUNE DENSITA’ (COMPLEMENTO ALLA LEZIONE PRECEDENTE)**

Riprendiamo l’esempio

X = seq(-5,5,0.01)

Y= dnorm(X)

plot(X,Y)

Si poteva automatizzare la scelta delle X ponendo:

X=rnorm(1000)

Y= dnorm(X)

plot(X,Y)

Aggiungiamo gli assi al grafico:

lines(c(-5,5),c(0,0),lwd=2)

lines(c(0,0),c(-0.1,0.5))

Analogamente, invece dei comandi:

Z = rnorm(10000)

hist(Z,50,FALSE)

X = seq(-5,5,0.01)

Y= dnorm(X,mean(Z),sd(Z))

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

si può automatizzare con:

Z0 = rnorm(10000); Z=sort(Z0)

hist(Z,50,FALSE)

Y= dnorm(Z,mean(Z),sd(Z))

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

Invece provare cosa succede senza il sort:

Z = rnorm(10000)

hist(Z,50,FALSE)

Y= dnorm(Z,mean(Z),sd(Z))

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

1. **CARICAMENTO DI VETTORI E TABELLE**

Supponiamo di avere un vettore di numeri scritto su un foglio come questo:

1.67 1.78 1.78 1.84 1.82 1.80

e supponiamo di volerlo caricare in R. Si opera così:

1. Scrivere (o copiare) su R il comando X <- scan("clipboard") senza dare d’invio.
2. Andare sul foglio dove c’è il vettore, selezionarlo e fare “copia”
3. Tornare col cursore su R e dare “invio”.

Verificare con “X invio” che i dati sono stati copiati.

Supponiamo di avere una tabella di numeri come questa:

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

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”.

**NOTA PER UTENTI MAC**: al sito http://stackoverflow.com/questions/9035674/r-function-to-copy-to-clipboard-on-mac-osx si trova che sostituendo a "clipboard" la stringa pipe("pbpaste") si ottiene lo stesso risultato, sia col comando scan che col comando read.table. Es. X <- scan(pipe("pbpaste")), A <- read.table (pipe("pbpaste")).

1. **REGRESSIONE LINEARE SEMPLICE**

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

Alt = A[,1]

Peso = A[,2]

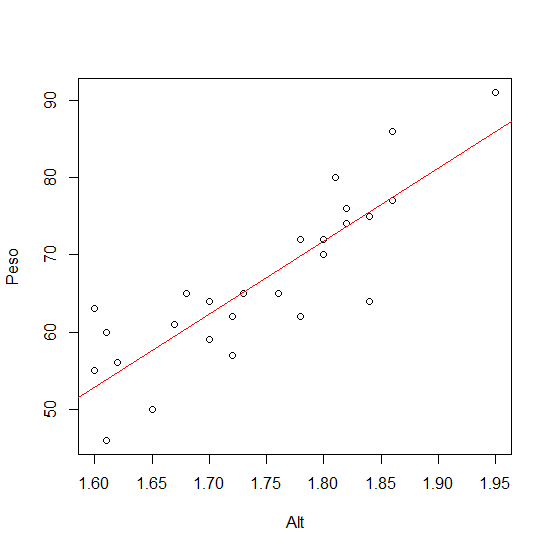
Data = A[,3]

Sesso = A[,4]

Ecco i comandi per visualizzare i punti nelle coordinate altezza-peso e la loro retta di regressione:

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.

Il comando abline aggiunge una retta ad un plot pre-esistente. La retta può essere specificata in più modi. Se i suoi coefficienti sono stati calcolati tramite una regressione lineare semplice, basta usare il comando come illustrato sopra. Se invece si conoscono intercetta s e coefficiente angolare m, il comando è

abline(s,m)

La regressione lineare semplice è stata eseguita dal comando lm(Peso~Alt). Esaminiamolo più a fondo: poniamo

REG = lm(Peso~Alt)

Possiamo interrogare l’esito del comando ad esempio chiedendo

summary(REG)

Ecco il risultato:

> summary(REG)

Call:

lm(formula = Peso ~ Alt)

Residuals:

Min 1Q Median 3Q Max

-11.5347 -2.5751 0.2492 2.2628 10.1684

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -98.52 19.37 -5.086 2.99e-05 \*\*\*

Alt 94.60 11.07 8.543 6.95e-09 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.404 on 25 degrees of freedom

Multiple R-squared: 0.7449, Adjusted R-squared: 0.7347

F-statistic: 72.98 on 1 and 25 DF, p-value: 6.95e-09

1. **CORRELAZIONE**

Possiamo calcolare il *coefficiente di correlazione* tra altezza e peso col comando

cor(Alt,Peso)

Il risultato è

> cor(Alt,Peso)

[1] 0.8630517

che è elevato. Si noti che il suo quadrato

cor(Alt,Peso)^2

> cor(Alt,Peso)^2

[1] 0.7448583

Coincide col numero “Multiple R-squared: 0.7449” del summary dela regressione.

NOTE.

1. Se non lo sappiamo già o non lo ritroviamo, come si trovano le istruzioni per tracciare una retta di regressione? Scrivendo su Google una frase ragionevole, tipo “add regression line with R”. Ecco cosa si impara:
2. abline(lm(Y~X),col="red",lwd=1.5) ispessisce un po’ la retta, se piace di più.
3. lines(lm(Y~X),col="red") non funziona.
4. Ci si può impratichire col comando abline, ad esempio eseguendo

abline(h=45)

abline(0, 1)