1. **SOLUZIONE 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?

**Soluzione di (i).** Calcoliamo la deviazione standard dei residui e prendiamo le rette che distano, in verticale, k\*sd dalla retta di regressione. La scelta di k è soggettiva. Può essere fatta direttamente o calcolata sulla base di una confidenza.

Vediamo un esempio. Preliminarmente rifacciamo una parte dell’esercitazione scorsa:

------- ------- ------- inizio richiamo ------- ------- -------

A <- 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

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]

cor(Alt,Peso)

plot(Alt,Peso)

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

------- ------- ------- fine richiamo ------- ------- -------

------- ------- ------- inizio aggiunte sulla regressione ------- ------- -------

REG = lm(Peso~Alt)

predict(REG) # oppure

fitted(REG) # forniscono i valori di y calcolati rispetto ai valori di x dei dati

residuals(REG) # fornisce i residui

coef(REG) # fornisce i coefficienti, iniziando dall’intercetta

------- ------- ------- fine aggiunte sulla regressione ------- ------- -------

------- ------- ------- soluzione esercizio ------- ------- -------

SD = sd(residuals(REG) )

plot(Alt,Peso)

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

abline(coef(REG) +c(2\*SD,0),col="green")

abline(coef(REG) -c(2\*SD,0),col="green")

Se si vuole calcolare k (di k\*sd) partendo da una confidenza, accettando di usare le gaussiane per la descrizione dei residui, cosa che un po’ possiamo controllare con

hist(residuals(REG) )

scegliamo una confidenza, es. 90% e calcoliamo

K = qnorm(0.95)

che fornisce il risultato 1.644854. Altro esempio: scegliendo una confidenza del 95%, calcoliamo

K = qnorm(0.975)

che è 1.96, circa il K=2 scelto sopra.

**Soluzione di (ii).** Seguiamo due strade, la prima parametrica, la seconda non parametrica.

------- ------- ------- inizio richiamo ------- ------- -------

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,FALSE)

------- ------- ------- fine richiamo ------- ------- -------

**Soluzione”parametrica” di (ii).** Immaginiamo che dietro l’istogramma ci sia una gaussiana.

Calcoliamo media e deviazione standard del campione, cioè di COR:

mean(COR); sd(COR)

disegnamo a titolo di verifica (e per gusto grafico) la gaussiana corrispondente

X = seq(-5,5,0.01)

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

lines(X,Y,col=3)

Ed infine calcoliamo l’area relativa al valore -0.41:

pnorm(-0.41,mean(COR), sd(COR))

Adottando un punto di vista bilaterale, possiamo concludere:

la probabilità di ottenere valori così estremi è

2\* pnorm(-0.41,mean(COR), sd(COR))

ovvero 0.037 (3,7%).

Siccome il fit col grafico gaussiano è estremamente accurato, il valore purtroppo è da accettare.

**Soluzione”non parametrica” di (ii).** Si tratta semplicemente di calcolare la percentuale di elementi del campione alla sinistra di -0.41.

Cerchiamo per tentativi:

C=sort(COR)

C[100]; C[200] … C[800]; C[850] = -0.410

Quindi ci sono circa 850/50000 elementi alla sinistra. Bilateralmente,

2\* 850/50000

ovvero 0.034, invece che 0.037.

1. **MATRICE DI CORRELAZIONE E PLOT DI UNA TABELLA**

cor(A)

plot(A)