1. **MIGLIORAMENTO DI UN COMANDO PRECEDENTE**

X = seq(-5,5,0.01)

Produce numeri a distanza 0.01 uno dall’altro, a partire da -5 fino a 5. Sostituisce quel complicato comando per fare grafici gaussiani. Tutte le esercitazioni precedenti sono state corrette.

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

A <- read.table ("clipboard")

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

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

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

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

cor(A)

plot(A)

1. **INDICATORI DI BENESSERE**

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

A <- 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

cor(A)

 PLIC SC SA.SC TD TMI

PLIC 1.0000000 0.3223197 -0.4110268 -0.3664348 -0.4432647

SC 0.3223197 1.0000000 -0.8417383 -0.8501667 -0.4834029

SA.SC -0.4110268 -0.8417383 1.0000000 0.9054169 0.5136762

TD -0.3664348 -0.8501667 0.9054169 1.0000000 0.4868433

TMI -0.4432647 -0.4834029 0.5136762 0.4868433 1.0000000

plot(A)



1. **PRIMO ESERCIZIO DI REGRESSIONE**

Costruire dei vettori con le colonne di A:

PLIC = A[,1]

SC = A[,2]

SA.SC = A[,3]

TD = A[,4]

TMI = A[,5]

Vogliamo un modello per TD basato sulle altre variabili. Eseguire la regressione

Reg1 = lm(TD~PLIC+SC+SA.SC+TMI)

Provare a chiedere Reg1. Chiedere poi summary(Reg1)

 summary(Reg1)

Call:

lm(formula = TD ~ PLIC + SC + SC.SA + TMI)

Residuals:

 Min 1Q Median 3Q Max

-0.82935 -0.20683 0.01375 0.23317 0.70087

Coefficients:

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

(Intercept) 0.0001030 0.0985596 0.001 0.99918

PLIC 0.0008371 0.1166154 0.007 0.99437

SC -0.3003680 0.1892197 -1.587 0.13327

SC.SA 0.6481730 0.1968985 3.292 0.00494 \*\*

TMI 0.0089889 0.1245278 0.072 0.94341

---

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

Residual standard error: 0.4408 on 15 degrees of freedom

Multiple R-squared: 0.8464, Adjusted R-squared: 0.8055

F-statistic: 20.67 on 4 and 15 DF, p-value: 5.801e-06

> sqrt(0.8464)

[1] 0.92

>

Osservazioni generiche: buon R^2, buon p-value globale (questo però in genere vale sempre), almeno un fattore importante e forse due; non è un cattivo modello. L’importanza dei fattori si vede dai relativi p-values.

Iniziamo un procedimento di eliminazione (si è discusso dell’importanza di ridurre e del significato di R^2 adjusted). Mai a coppie o gruppi, come chiarito anche dalla teoria dei fattori allineati.

 Osservando i p-value dei singoli fattori e la correlazione tra TD e le altre variabili, è chiaro che la peggiore è PLIC, seguita da TMI. Eseguiamo la regressione senza PLIC:

Reg2 = lm(TD~SC+SA.SC+TMI)

Poi chiediamo summary(Reg2)

> summary(Reg2)

Call:

lm(formula = TD ~ SC + SC.SA + TMI)

Residuals:

 Min 1Q Median 3Q Max

-0.8290 -0.2070 0.0137 0.2329 0.7014

Coefficients:

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

(Intercept) 0.000103 0.095430 0.001 0.9992

SC -0.300484 0.182547 -1.646 0.1192

SC.SA 0.647871 0.186253 3.478 0.0031 \*\*

TMI 0.008717 0.114857 0.076 0.9404

---

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

Residual standard error: 0.4268 on 16 degrees of freedom

Multiple R-squared: 0.8464, Adjusted R-squared: 0.8176

F-statistic: 29.4 on 3 and 16 DF, p-value: 9.597e-07

>

R^2 non è diminuito. R^2 adjusted è migliorato. Persino i singoli p-values sono migliorati. Il modello è sicuramente migliore.

TMI può essere ulteriormente eliminato, non era un fattore allineato (un poco, ma è irrilevante).