

Linear Regression

Consider the wine data. The data come from a study of Pinot Noir wine quality. The dataset contains 38 observations and 7 variables: Quality, Clarity, Aroma, Body, Flavor, Oakiness, and Region. The goal is to develop a model that relates the quality of Pinot Noir with its features. The model can potentially be used to predict the quality of the wine.

- a) **Figure** represent the Scatterplot matrix for wine data. There is a strong positive correlation between response variable Quality and predictor variables Flavor and Aroma, moderate correlation between Quality and Body. Moreover, there is a strong positive correlation between predictor variables Aroma and Flavor and Body and Flavor.

```
library(car)
library(lmtest)
library(ggplot2)

wine<-read.table("wine.txt",header = TRUE)
#View(wine)
wine$Region<-as.factor(wine$Region)
str(wine)

'data.frame':  38 obs. of  7 variables:
 $ Clarity : num  1 1 1 1 1 1 1 1 1 1 ...
 $ Aroma   : num  3.3 4.4 3.9 3.9 5.6 4.6 4.8 5.3 4.3 4.3 ...
 $ Body    : num  2.8 4.9 5.3 2.6 5.1 4.7 4.8 4.5 4.3 3.9 ...
 $ Flavor  : num  3.1 3.5 4.8 3.1 5.5 5 4.8 4.3 3.9 4.7 ...
 $ Oakiness: num  4.1 3.9 4.7 3.6 5.1 4.1 3.3 5.2 2.9 3.9 ...
 $ Quality : num  9.8 12.6 11.9 11.1 13.3 12.8 12.8 12 13.6 13.9 ...
 $ Region  : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 3 1 ...

cor(wine[1:6])

      Clarity      Aroma      Body      Flavor      Oakiness      Quality
Clarity  1.00000000  0.0619021 -0.3083783 -0.08515993  0.18321471  0.02844131
Aroma    0.06190210  1.0000000  0.5489102  0.73656121  0.20164445  0.70732432
Body     -0.30837826  0.5489102  1.0000000  0.64665917  0.15210591  0.54870219
Flavor   -0.08515993  0.7365612  0.6466592  1.00000000  0.17976051  0.79004713
Oakiness  0.18321471  0.2016444  0.1521059  0.17976051  1.00000000 -0.04704047
Quality  0.02844131  0.7073243  0.5487022  0.79004713 -0.04704047  1.00000000

panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...) {
  usr <- par("usr")
  on.exit(par(usr))
  par(usr = c(0, 1, 0, 1))
  r <- abs(cor(x, y, use = "complete.obs"))
  txt <- format(c(r, 0.123456789), digits = digits)[1]
  txt <- paste(prefix, txt, sep = " ")
  if (missing(cex.cor)) cex.cor <- 0.8/strwidth(txt)
  text(0.5, 0.5, txt, cex = cex.cor * (1 + r) / 2)
}

panel.hist <- function(x, ...) {
  usr <- par("usr")
  on.exit(par(usr))
  par(usr = c(usr[1:2], 0, 1.5) )
```

```

h <- hist(x, plot = FALSE)
breaks <- h$breaks
nB <- length(breaks)
y <- h$counts
y <- y/max(y)
rect(breaks[-nB], 0, breaks[-1], y, ...)
}

#Scatter plot matrix for wine data
my_cols <- c("#00AFBB", "#E7B800", "#FC4E07")
pairs(wine[,1:6], col = my_cols[wine$Region], pch=20, upper.panel = panel.cor, diag.panel = panel.hist, oma=c(2,2,2,2))

```

