MA575 Lab6

Shuqiang Chen

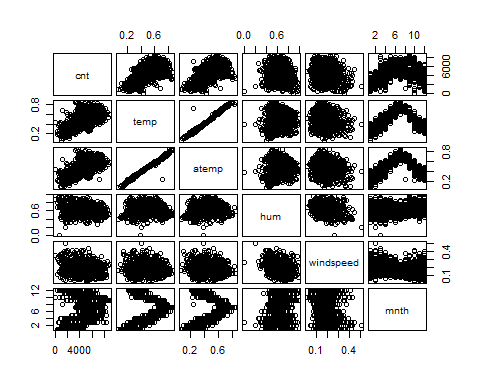
10.24.2018

# 1. Choose the response variable (Y) and one covariate (X).   
#Please put some though for your response and covariate variable selection.  
setwd("C:/Users/daish/Desktop/Lab 6")  
# import dataset day.csv  
BikeSharingInDay <- read.csv(file = "day.csv")  
# Sneak peak at the data  
head(BikeSharingInDay)

## instant dteday season yr mnth holiday weekday workingday weathersit  
## 1 1 2011/1/1 1 0 1 0 6 0 2  
## 2 2 2011/1/2 1 0 1 0 0 0 2  
## 3 3 2011/1/3 1 0 1 0 1 1 1  
## 4 4 2011/1/4 1 0 1 0 2 1 1  
## 5 5 2011/1/5 1 0 1 0 3 1 1  
## 6 6 2011/1/6 1 0 1 0 4 1 1  
## temp atemp hum windspeed casual registered cnt  
## 1 0.344167 0.363625 0.805833 0.1604460 331 654 985  
## 2 0.363478 0.353739 0.696087 0.2485390 131 670 801  
## 3 0.196364 0.189405 0.437273 0.2483090 120 1229 1349  
## 4 0.200000 0.212122 0.590435 0.1602960 108 1454 1562  
## 5 0.226957 0.229270 0.436957 0.1869000 82 1518 1600  
## 6 0.204348 0.233209 0.518261 0.0895652 88 1518 1606

# choose counts of total rental bikes as response variable (Y)  
# hypothesis: temp(X1),atemp(X2),hum(X3),windspeed(X4),mnth(X5) have impact on counts of total rental bikes(Y)  
cnt <- BikeSharingInDay$cnt  
  
# choose covariate (X1~X5)  
temp <- BikeSharingInDay$temp  
atemp <- BikeSharingInDay$atemp  
hum <- BikeSharingInDay$hum  
windspeed <- BikeSharingInDay$windspeed  
mnth <- BikeSharingInDay$mnth

# 2. Plot Y VS. X1-x5 (i.e. a scatterplot) from the data.  
# Plot scatter matrix   
pairs(~cnt+temp+atemp+hum+windspeed+mnth,gap=0.4)



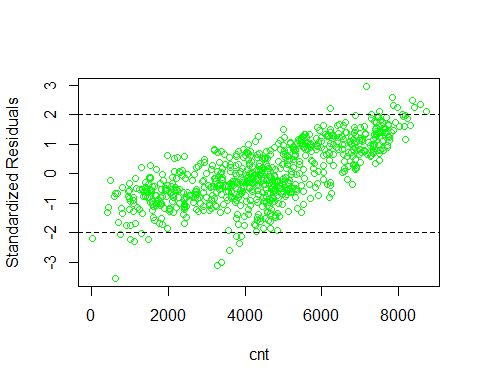
m.mls <- lm(cnt ~ temp + atemp + hum + windspeed + mnth)  
summary(m.mls)

##   
## Call:  
## lm(formula = cnt ~ temp + atemp + hum + windspeed + mnth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4829.6 -992.7 -188.5 1089.4 3615.1   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3542.46 350.99 10.093 < 2e-16 \*\*\*  
## temp 2133.18 2228.44 0.957 0.3388   
## atemp 4752.92 2517.11 1.888 0.0594 .   
## hum -3526.16 380.11 -9.277 < 2e-16 \*\*\*  
## windspeed -3963.63 710.29 -5.580 3.39e-08 \*\*\*  
## mnth 95.04 15.74 6.037 2.50e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1389 on 725 degrees of freedom  
## Multiple R-squared: 0.4895, Adjusted R-squared: 0.486   
## F-statistic: 139 on 5 and 725 DF, p-value: < 2.2e-16

#Perform MLR using R on your response (Y) and covariates (X1,X2, ..Xr) .  
  
m.mls <- lm(cnt ~ atemp + hum + windspeed + mnth)  
summary(m.mls)

##   
## Call:  
## lm(formula = cnt ~ atemp + hum + windspeed + mnth)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4864.6 -993.4 -171.3 1102.3 4114.5   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3455.26 338.94 10.194 < 2e-16 \*\*\*  
## atemp 7141.87 328.01 21.774 < 2e-16 \*\*\*  
## hum -3544.62 379.60 -9.338 < 2e-16 \*\*\*  
## windspeed -3845.58 699.46 -5.498 5.33e-08 \*\*\*  
## mnth 95.02 15.74 6.036 2.51e-09 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1389 on 726 degrees of freedom  
## Multiple R-squared: 0.4889, Adjusted R-squared: 0.486   
## F-statistic: 173.6 on 4 and 726 DF, p-value: < 2.2e-16

# Standard Residuals  
StanResMLS <- rstandard(m.mls)  
par(mfrow=c(1,1))  
plot(cnt,StanResMLS,xlab="cnt", ylab="Standardized Residuals", col="green")  
abline(h=2,lty=2)  
abline(h=-2,lty=2)



X <- cbind(cnt, atemp, hum, windspeed, mnth)  
c <- cor(X)  
round(c,3)

## cnt atemp hum windspeed mnth  
## cnt 1.000 0.631 -0.101 -0.235 0.280  
## atemp 0.631 1.000 0.140 -0.184 0.227  
## hum -0.101 0.140 1.000 -0.248 0.222  
## windspeed -0.235 -0.184 -0.248 1.000 -0.208  
## mnth 0.280 0.227 0.222 -0.208 1.000