Midterm

Yishan Liu

November 7, 2015

# Title about the paper

## Step 1: Read the data

setwd("/Users/Gracie/Desktop/BigData")  
mydata<-read.csv("ozone\_data.csv")

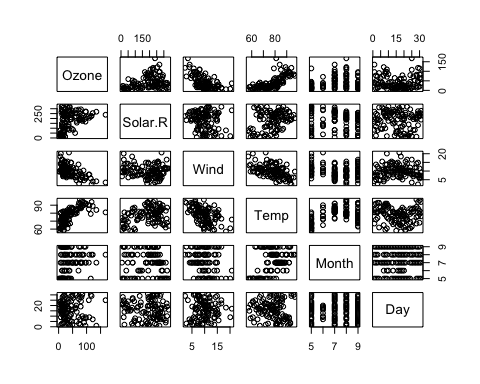
## Step 2: linear regression

model<-lm(Ozone~Temp+Wind,data=mydata)  
summary(model)

##   
## Call:  
## lm(formula = Ozone ~ Temp + Wind, data = mydata)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -42.156 -13.216 -3.123 10.598 98.492   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -67.3220 23.6210 -2.850 0.00524 \*\*   
## Temp 1.8276 0.2506 7.294 5.29e-11 \*\*\*  
## Wind -3.2948 0.6711 -4.909 3.26e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 21.73 on 108 degrees of freedom  
## Multiple R-squared: 0.5814, Adjusted R-squared: 0.5736   
## F-statistic: 74.99 on 2 and 108 DF, p-value: < 2.2e-16

Summary:1.model 2.R square ##Step 3: Correlation Plot and Pearson's correlation coefficient

pairs(mydata)



cor(mydata)

## Ozone Solar.R Wind Temp Month  
## Ozone 1.000000000 0.34834169 -0.61249658 0.6985414 0.142885168  
## Solar.R 0.348341693 1.00000000 -0.12718345 0.2940876 -0.074066683  
## Wind -0.612496576 -0.12718345 1.00000000 -0.4971897 -0.194495804  
## Temp 0.698541410 0.29408764 -0.49718972 1.0000000 0.403971709  
## Month 0.142885168 -0.07406668 -0.19449580 0.4039717 1.000000000  
## Day -0.005189769 -0.05775380 0.04987102 -0.0965458 -0.009001079  
## Day  
## Ozone -0.005189769  
## Solar.R -0.057753801  
## Wind 0.049871017  
## Temp -0.096545800  
## Month -0.009001079  
## Day 1.000000000

Analyze the plot and pearson cc, try new model and see the result ##Step 4:Confidence interval

confint(model,level=0.95)

## 2.5 % 97.5 %  
## (Intercept) -114.142962 -20.500944  
## Temp 1.330917 2.324192  
## Wind -4.625171 -1.964507