Lab 4 Ordinary Least Squares

This is an R Markdown Notebook. When you execute code within the notebook, the results appear beneath the code.

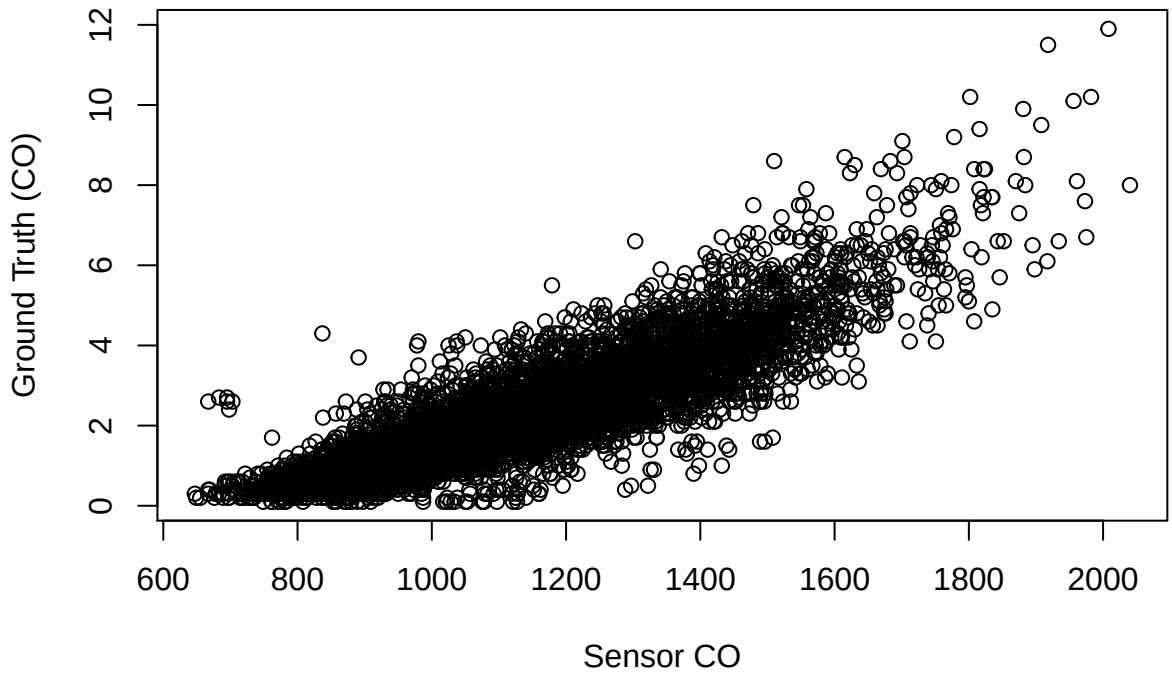
Try executing this chunk by clicking the *Run* button within the chunk or by placing your cursor inside it and pressing *Ctrl+Shift+Enter*.

```
# Read data from cvs file with ";" instead of ","
pollutiondata <- read.csv2("AirQualityData.csv",header=TRUE)

# Replace missing data i.e. -200 with NA
pollutiondata[pollutiondata==-200]<-NA
```

```
# Force Temperature and pollutant to be a numeric vector
Temperature = as.numeric(as.character(pollutiondata$Temperature))
RelativeHumidity = as.numeric(as.character(pollutiondata$RelativeHumidity))
SensorCO = as.numeric(as.character(pollutiondata$PT08.S1.CO.))
GroundCO = as.numeric(as.character(pollutiondata$CO.GT.))
GroundCO[GroundCO == -200]<-NA
```

```
# Create temporary data set with Temperature and Carbon Monoxide pollutant
tempdataset <- data.frame(SensorCO, GroundCO, Temperature, RelativeHumidity)
rm(Temperature,SensorCO,GroundCO,RelativeHumidity)

# Remove rows with NA
Dataset<-tempdataset[complete.cases(tempdataset), ]
attach(Dataset)
```

```
# Scatterplot
plot(SensorCO,GroundCO,ylab="Ground Truth (CO)", xlab="Sensor CO")
```



```r
# Ordinary LS
m.ols <- lm(GroundCO~SensorCO)
summary(m.ols)
```

```
##
## Call:
## lm(formula = GroundCO ~ SensorCO)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9862 -0.3917 -0.0342  0.3206  4.7067
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.285e+00  4.133e-02  -103.7   <2e-16 ***
## SensorCO     5.776e-03  3.651e-05   158.2   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6842 on 7342 degrees of freedom
## Multiple R-squared:  0.7731, Adjusted R-squared:  0.7731
## F-statistic: 2.502e+04 on 1 and 7342 DF,  p-value: < 2.2e-16
```

```r
#95% confidence intervals of OLS
round(confint(m.ols,level=0.95),6)
```

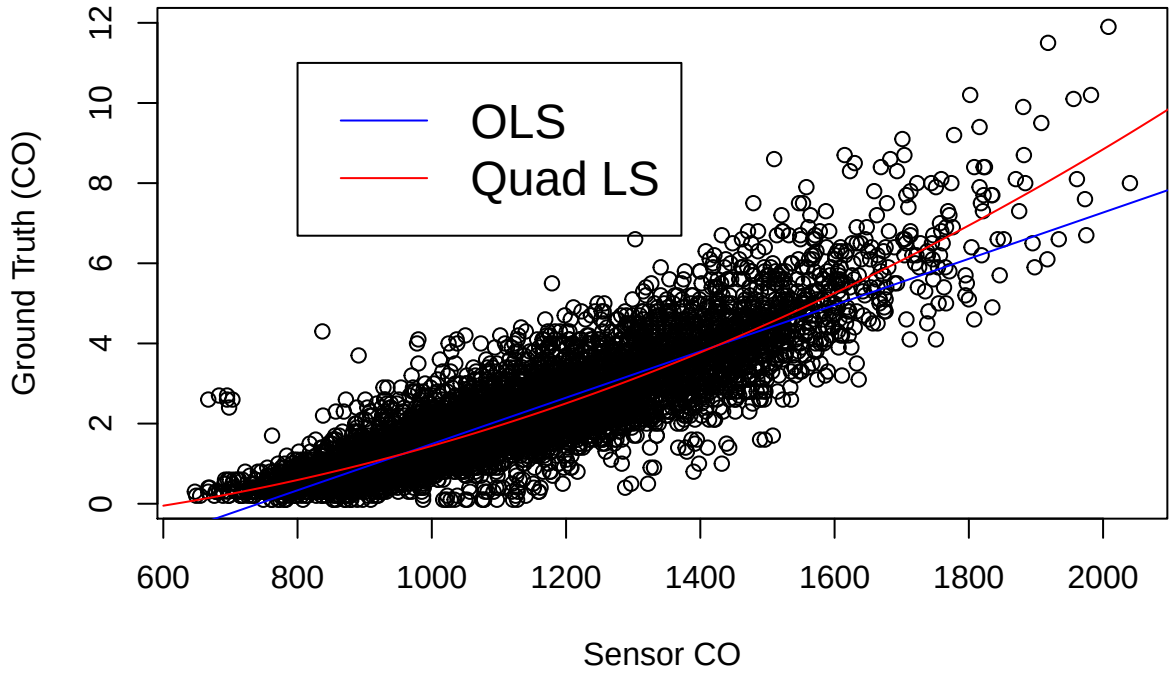
```
##                 2.5 %    97.5 %
## (Intercept) -4.365865 -4.203830
## SensorCO     0.005704  0.005847
```

```r
# Quadratic LS
m.quadls <- lm(GroundCO~SensorCO + I(SensorCO^2))
summary(m.quadls)
```

```
##
## Call:
## lm(formula = GroundCO ~ SensorCO + I(SensorCO^2))
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.9977 -0.3653 -0.0478  0.2939  4.0370
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -7.159e-01  1.709e-01  -4.189 2.84e-05 ***
## SensorCO      -4.603e-04  2.925e-04  -1.574    0.116
## I(SensorCO^2)  2.620e-06  1.220e-07  21.482  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6637 on 7341 degrees of freedom
## Multiple R-squared:  0.7866, Adjusted R-squared:  0.7865
## F-statistic: 1.353e+04 on 2 and 7341 DF,  p-value: < 2.2e-16
```

```r
# Scatter plot
plot(SensorCO,GroundCO,ylab="Ground Truth (CO)", xlab="Sensor CO")
```

```r
abline(lsfit(SensorCO,GroundCO),col="blue")
SensorCONew<-seq(600,2200,len=100)
lines(SensorCONew,predict(m.quadls,newdata=data.frame(SensorCO=SensorCONew)),col="red")
legend(800, 11, legend=c("OLS", "Quad LS"), col=c("blue", "red"), lty=1, cex=1.5)
```




```r
detach(Dataset)

# Exercise: Perform linear regression on your project data. Choose your response and covariate.
```