

Applied Statistics Math 564 HW2 Solution :5

Arinjay Jain (A20447307) ajain80@hawk.iit.edu

```
#random 100 row from given data.
df <- APPENC07[sample(nrow(APPENC07), 100), ]
#sort data in ascending order
sortDataFrame <- df[order(df$V3), ]

# assigning second column sales price (Y)
Ygiven <- sortDataFrame$V2
# assigning 3rd column finished square feet (X)
Xgiven <- sortDataFrame$V3

# L is sum((Xgiven-mean(Xgiven))^2)
L <- sum((Xgiven - mean(Xgiven)) ^ 2)

#the point of estimate b1 for B1
b1 <- sum((Xgiven - mean(Xgiven)) * (Ygiven - mean(Ygiven))) / L
c(b1)= 147.5711
#the point of estimate bo for BO
bo <- mean(Ygiven) - b1 * mean(Xgiven)
c(bo)= -58545.81
#estimated Yhat of response variable y
yhat <- bo + b1 * Xgiven

plot(Xgiven, Ygiven)
lines(Xgiven, yhat, col = "red", lwd = "2")
abline(h = 0, v = 0)
#Dof n-2
n <- 100
# t value as (1-alpha/2 = 0.975) and (n-2 = 98)DOF
t <- qt(0.95, n - 2)
c(t)= 1.660551

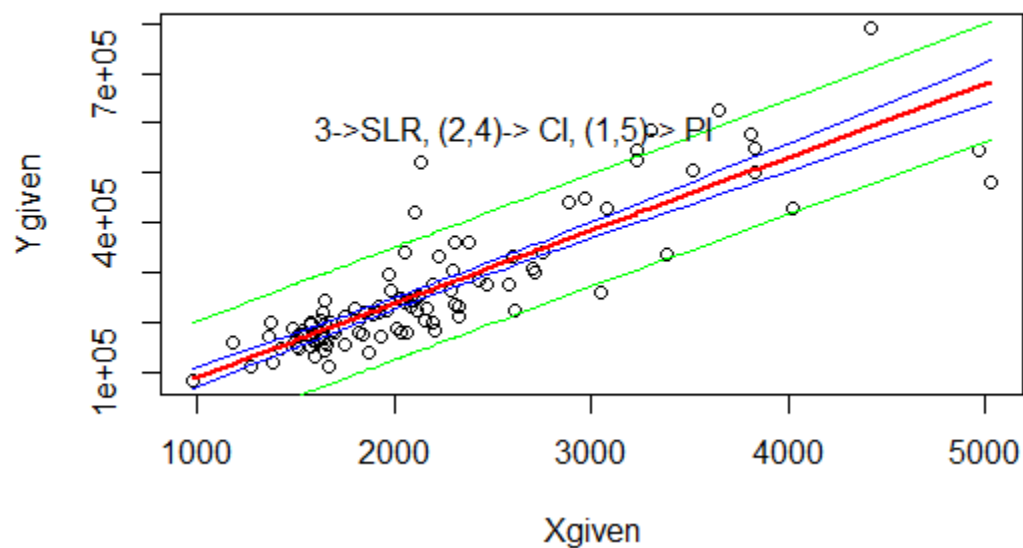
#MSE
MSE <- sum((Ygiven - yhat)^2) / n - 2
c(MSE) = 4580674619

#Standard error (Se) of yhat for confident interval
for (i in 1:length(Xgiven)) {
  #Some part of Se, G=(xh-xbar)^2/sum((Xgiven-xbar)^2)
  G[i] <- ((Xgiven[i] - mean(Xgiven)) ^ 2) / L
  # Se of mean response yh
  Se[i] <- sqrt(MSE * (1 / n + G[i]))
  sePI[i] <- sqrt(MSE * (1 + 1 / n + G[i]))
  #yhat H (yh) for every value of Xh (Xgiven[i])
  yh[i] <- bo + b1 * Xgiven[i]
  #confident Interval of mean response.
```

Applied Statistics Math 564 HW2 Solution :5

Arinjay Jain (A20447307) ajain80@hawk.iit.edu

```
Clpos <- yh + t * Se
Clneg <- yh - t * Se
Plpos <- yh + t * sePI
Plneg <- yh - t * sePI
}
lines(Xgiven, Clpos, col = 'blue')
lines(Xgiven, Clneg, col = 'blue')
lines(Xgiven, Plpos, col = 'green')
lines(Xgiven, Plneg, col = 'green')
#text(locator(), labels = c("blue lines is CI", "green lines is PI"))
text(x=Xgiven[80], y=Ygiven[90], label="3->SLR, (2,4)-> CI, (1,5)-> PI")
```



#part-2

#CI and PI at X=xbar mean

```
mean(Ygiven)=265772.7
Clmeanpos<-mean(Ygiven)+t*sqrt(MSE * (1 / n ))
c(Clmeanpos)= 277011.4
Clmeanneg<-mean(Ygiven)-t*sqrt(MSE * (1 / n ))
c(Clmeanneg)= 254534
Plmeanpos<-mean(Ygiven)+t*sqrt(MSE * (1+1 / n ))
c(Plmeanpos)= 378720.5
Plmeanneg<-mean(Ygiven)-t*sqrt(MSE * (1+1 / n ))
c(Plmeanneg)= 152824.9
```