# Homework 7

*Joshua Oswari - A14751270*

*05/22/2019*

Problem 1

Part (a)

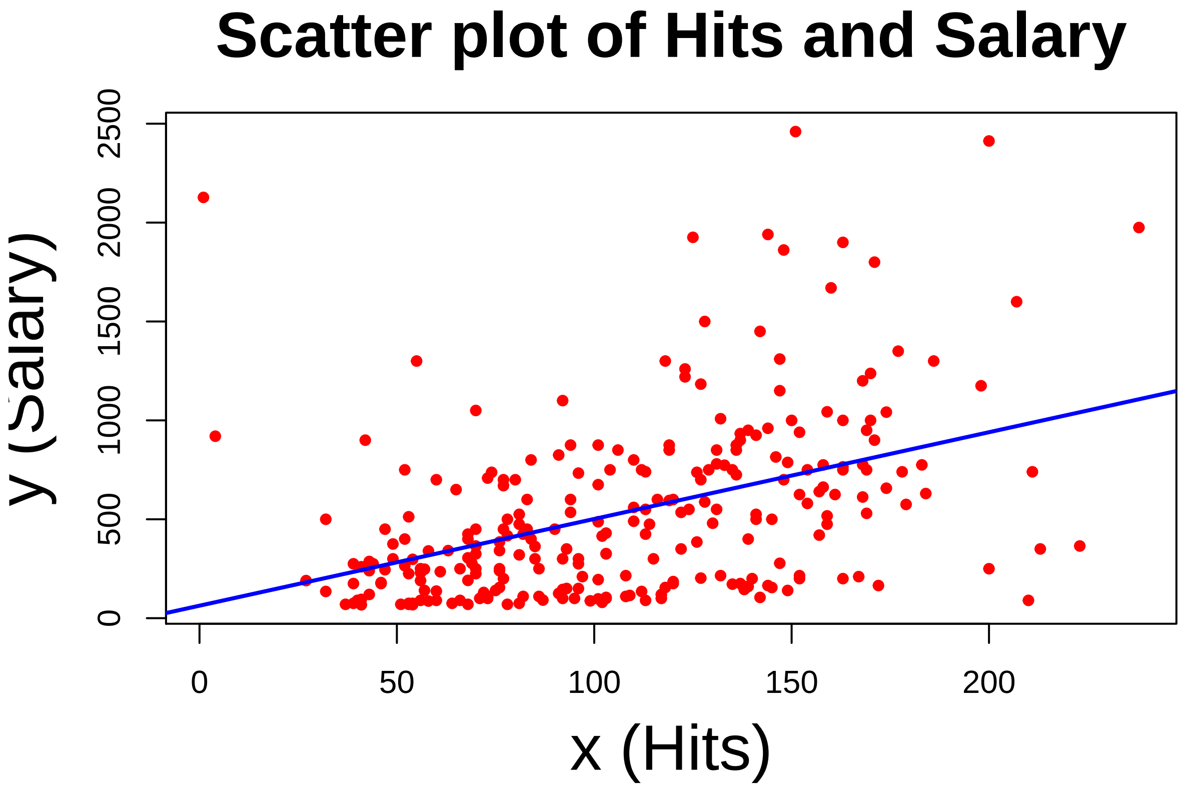
```
data = read.csv("~/Documents/Math189/HW7/baseball_5.csv",header = T)

#draw a scatterplot
plot(x = data$Hits, y = data$Salary,
 xlab = "x (Hits)", ylab = "y (Salary)",
 main = "Scatter plot of Hits and Salary",
 col="red", pch=20, cex.lab=2, cex.main=2
)

slm.fit =lm(Salary~Hits ,data=data)
summary(slm.fit)
```

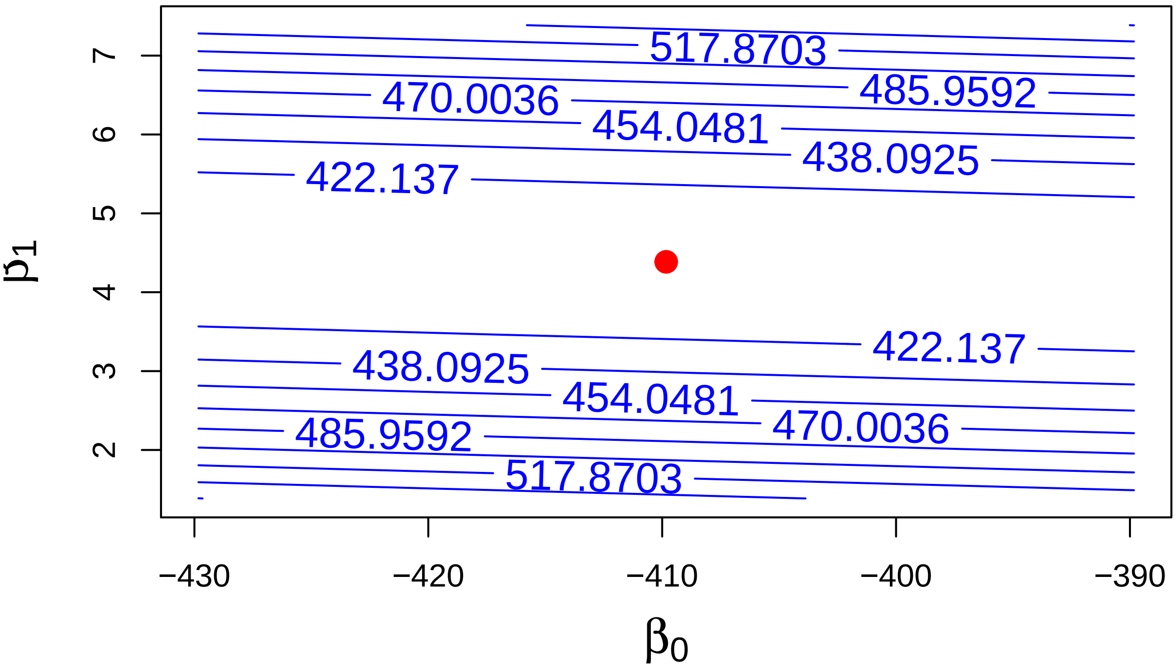
```
##
## Call:
## lm(formula = Salary ~ Hits, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -893.99 -245.63  -59.08  181.12 2059.90
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  63.0488    64.9822   0.970    0.333
## Hits          4.3854     0.5561   7.886 8.53e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 406.2 on 261 degrees of freedom
## Multiple R-squared:  0.1924, Adjusted R-squared:  0.1893
## F-statistic: 62.19 on 1 and 261 DF,  p-value: 8.531e-14
```

```
###### Simple Linear Regression Fit ######
plot(x = data$Hits, y = data$Salary,
 xlab = "x (Hits)", ylab = "y (Salary)",
 main = "Scatter plot of Hits and Salary",
 col="red", pch=20, cex.lab=2, cex.main=2
)
abline(slm.fit, col="blue", lwd=2)
```



```
g=50
x=data$Hits
y=data$Salary
n=length(y)
b=sum((x-mean(x))*(y-mean(y)))/sum((x-mean(x))^2)
a=mean(y)-b*mean(x)
RSS.min=sum((y-as.vector(cbind(1,x)%*%c(a,b)))^2)/(n-2)
a.grid=seq(a-20,a+20,length=g)
b.grid=seq(b-3,b+3,length=g)
grid=as.matrix(expand.grid(a.grid,b.grid))
RSS=rep(0,g^2)
for (i in 1:(g^2)){
 yhat=as.vector(cbind(1,x)%*%grid[i,])
 RSS[i]=sum((y-yhat)^2)/(n-2)
}
RSE=sqrt(RSS)
RSS=matrix(RSS,g,g)
RSE=matrix(RSE,g,g)
m=which.min(RSE)

#plot RSS
contour(a.grid-b*mean(data$Hits),b.grid,RSE,xlab=expression(beta[0]),ylab=expression(beta[
1]),levels=seq(min(RSE), max(RSE),
length.out=10),axes=T,frame.plot=T,col=4,drawlabels=T,cex.lab=1.5,labcex=1.3)
points(a-b*mean(data$Hits),b,col=2,pch=19,cex=1.5)
```



Part 2

```
mlm.fit =lm(Walks~Hits+PutOuts+CHits,data=data)

summary(mlm.fit)
```

```
##
## Call:
## lm(formula = Walks ~ Hits + PutOuts + CHits, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -43.227 -11.846  -2.161  10.442  57.597
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.141820   2.849852   2.857  0.00462 **
## Hits        0.250032   0.025428   9.833  < 2e-16 ***
## PutOuts     0.008976   0.003993   2.248  0.02541 *
## CHits       0.004711   0.001693   2.783  0.00578 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.26 on 259 degrees of freedom
## Multiple R-squared:  0.3757, Adjusted R-squared:  0.3685
## F-statistic: 51.95 on 3 and 259 DF,  p-value: < 2.2e-16
```

```
RSS0=19473
RSS1=15418
p0=1
p=3
n=506
F=(RSS0-RSS1)*(n-p-1)/RSS1/(p-p0)
pf(F, p-p0, n-p-1, lower.tail=F)
```

```
## [1] 1
```

```
library(leaps)
regfit.full=regsubsets (CHits~., data=data) #
summary(regfit.full)
```

```
## Subset selection object
## Call: regsubsets.formula(CHits ~ ., data = data)
## 4 Variables  (and intercept)
##         Forced in Forced out
## Salary      FALSE      FALSE
## Hits        FALSE      FALSE
## Walks       FALSE      FALSE
## PutOuts     FALSE      FALSE
## 1 subsets of each size up to 4
## Selection Algorithm: exhaustive
##          Salary Hits Walks PutOuts
## 1  ( 1 ) "*"    " "  " "   " "
## 2  ( 1 ) "*"    " "  " "   "*"
## 3  ( 1 ) "*"    " "  "*"   "*"
## 4  ( 1 ) "*"    "*"  "*"   "*"
```

Part 3

```
# In multivariate regression there are more than one dependent
# variable with different variances (or distributions).
#The predictor variables may be more than one or multiple.
#So it is may be a multiple regression with a matrix of dependent
# variables, i. e. multiple variances.
# in conclusion, the simple linear fit this data the best
```