Math 189 HW 7

Group Members:

Yaqi Chen (PID: A15742547; Section: A03)

Yuetong Lyu (PID: A 13779993; Section: A04)

Siyi He (PID: A13400569; Section: A03)

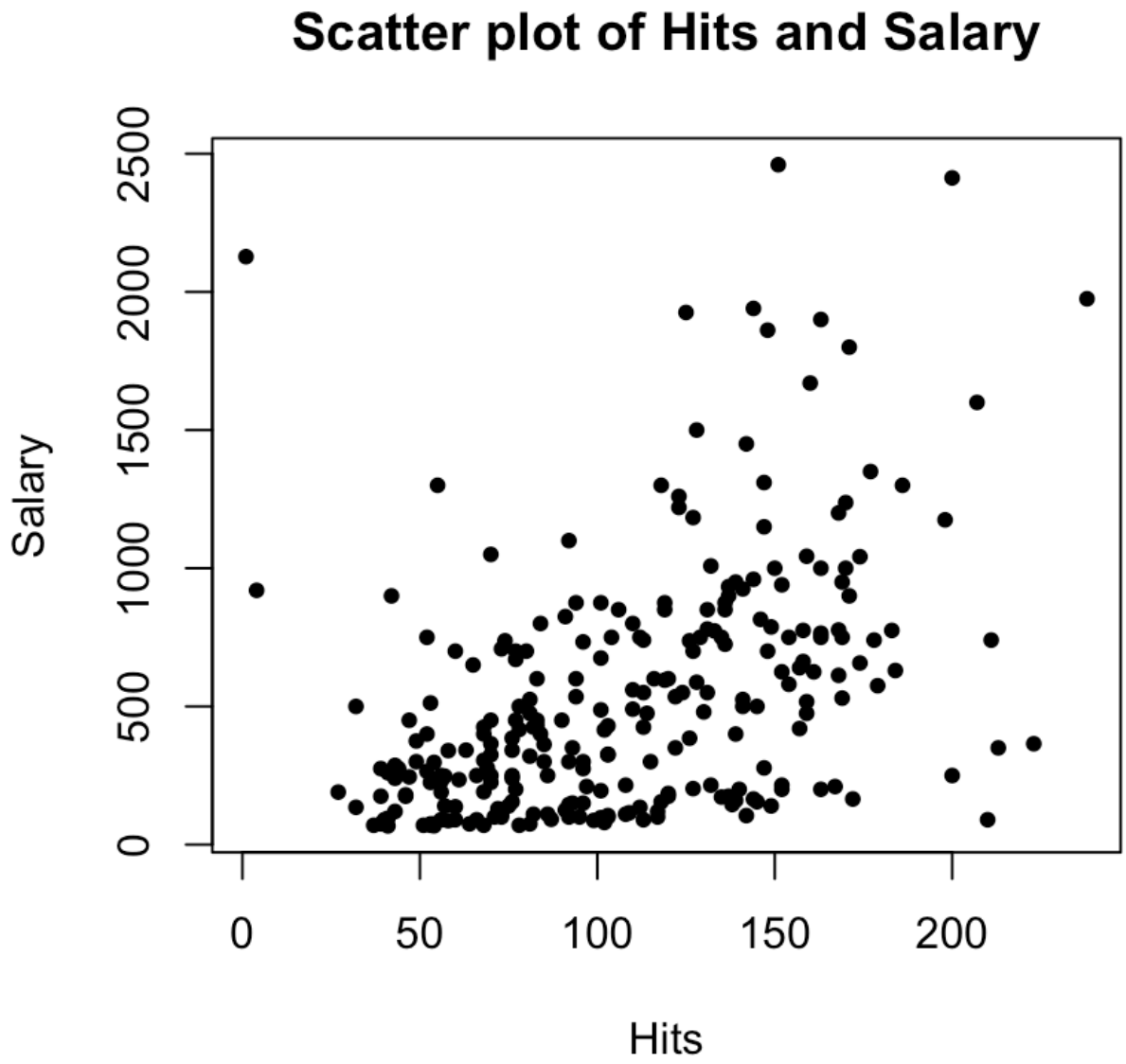
Zhenyuan Xu (PID: A92067995; Section: A02)

Jiawei Chao (PID: A138001; Section: A02)

Problem 1A

```
#Draw a scatter plot between Hits and Salary.
plot(x = data$Hits, y = data$Salary,
     xlab = "Hits", ylab = "Salary" ,
     main = "Scatter plot of Hits and Salary",
     col=1, pch=20)

#Consider a simple linear regression using Hits as predictor. Estimate
#the regression coefficients and their standard errors
n=dim(data)[1]
y=data$Salary
x=data$Hits
fit=lm(y~x)
abline(fit)
summary(fit)

coefficient1=fit$coefficients[2]
coefficient0=fit$coefficients[1]

RSS=sum((y-coefficient0-coefficient1*x)^2)
TSS=sum((y-mean(y))^2)
R2=(TSS-RSS)/TSS
```

Output:



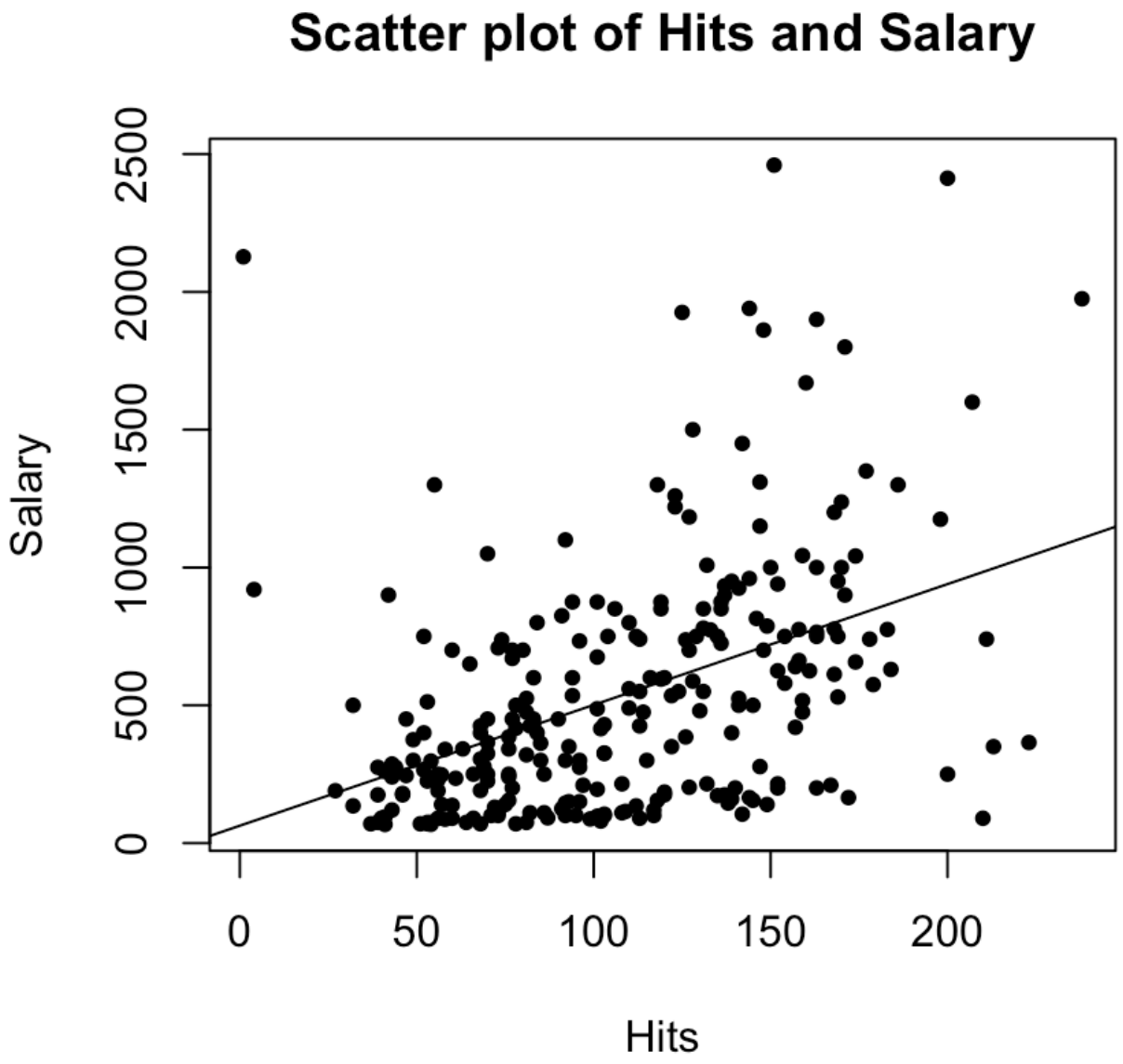

```
Call:
lm(formula = y ~ x)

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
x             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
```

the Residual standard error is 406.2

the Intercept coefficient is 63.0488, standard error is 64.9822

the Hits coefficient is 4.3854, standard error is 0.5561



I think this line doesn't fit the data well.

```
> RSS
[1] 43058621
> R2
[1] 0.1924355
```

RSS is 43058621

the square of R is 0.1924355

Problem 1B

```
#Consider a multivariate linear model using Hits, Walks, PutOuts and CHits as
predictors.
multifit=lm(data$Salary~data$Hits+data$Walks+data$PutOuts+data$CHits)
summary(multifit)

coefficient_0=multifit$coefficients[1]
coefficient_Hits=multifit$coefficients[2]
coefficient_Walks=multifit$coefficients[3]
coefficient_Putouts=multifit$coefficients[4]
coefficient_Chits=multifit$coefficients[5]
RSS_multi=sum((y-coefficient_0-coefficient_Hits*data$Hits-
coefficient_Walks*data$Walks-coefficient_Putouts*data$PutOuts
               -coefficient_Chits*data$CHits )^2)
R2_multi=(TSS-RSS_multi)/TSS
```

Output:

```
Call:
lm(formula = data$Salary ~ data$Hits + data$Walks + data$PutOuts +
    data$CHits)

Residuals:
    Min      1Q  Median      3Q     Max
-811.49 -169.57  -40.38  108.18 2211.38

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -109.83481   56.44049  -1.946 0.052737 .
data$Hits       1.84601    0.58106   3.177 0.001669 **
data$Walks      3.46111    1.21166   2.857 0.004632 **
```

```
data$PutOuts    0.27091    0.07861   3.446 0.000664 ***
data$CHits      0.31246    0.03350   9.328  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 336.6 on 258 degrees of freedom
Multiple R-squared:  0.4519,  Adjusted R-squared:  0.4434
F-statistic: 53.18 on 4 and 258 DF,  p-value: < 2.2e-16
```

the Intercept coefficient is -109.83481 with standard error 56.44049

the Hits coefficient is 1.84601 with standard error 0.58106

the Walks coefficient is 3.46111 with standard error 1.21166

the PutOuts coefficient is 0.27091 with standard error 0.07861

the CHits coefficient is 031246 with standard error 0.03350

```
> RSS_multi
[1] 29223384
> R2_multi
[1] 0.4519154
```

RSS is 29223384

the square of R is 0.4519154

The p-values of all coefficients are less than 0.05, so we reject all testing results.

Problem 1C

```
p0=1
p=4
F=(RSS-RSS_multi)*(n-p-1)/RSS_multi/(p-p0)
pf(F, p-p0, n-p-1, lower.tail=FALSE)
```

Output:

```
> pf(F, p-p0, n-p-1, lower.tail=FALSE)
[1] 1.417223e-21
```

the p-value is 1.4177223e-21 which is very small, therefore we reject the null model.