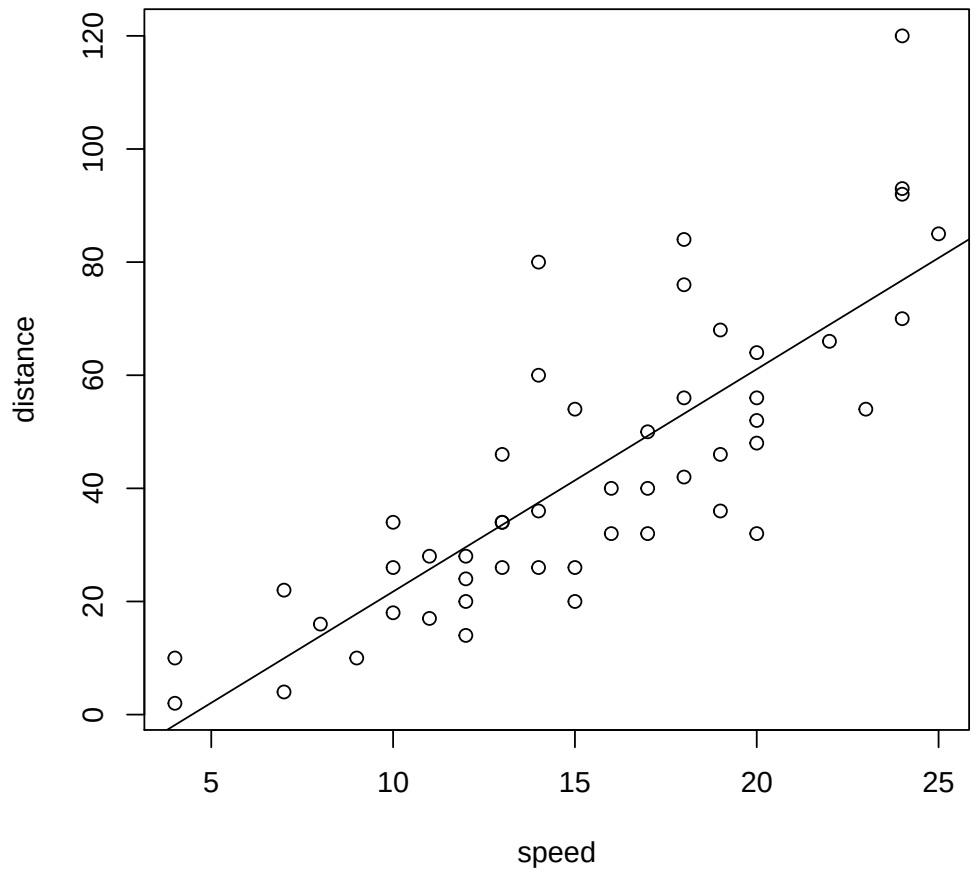
Homework 5 - STAT 511

Amal Agarwal

Answer 5

```
> # extracting the cars dataset
> data(cars)
> # response (speed) and predictor (distance) vectors
> speed<-cars$speed
> distance<-cars$dist
> # fitting the simple linear model
> fit=lm(distance~speed)
> # scatter plot with fitted line
> plot(speed,distance)
> abline(fit)
```



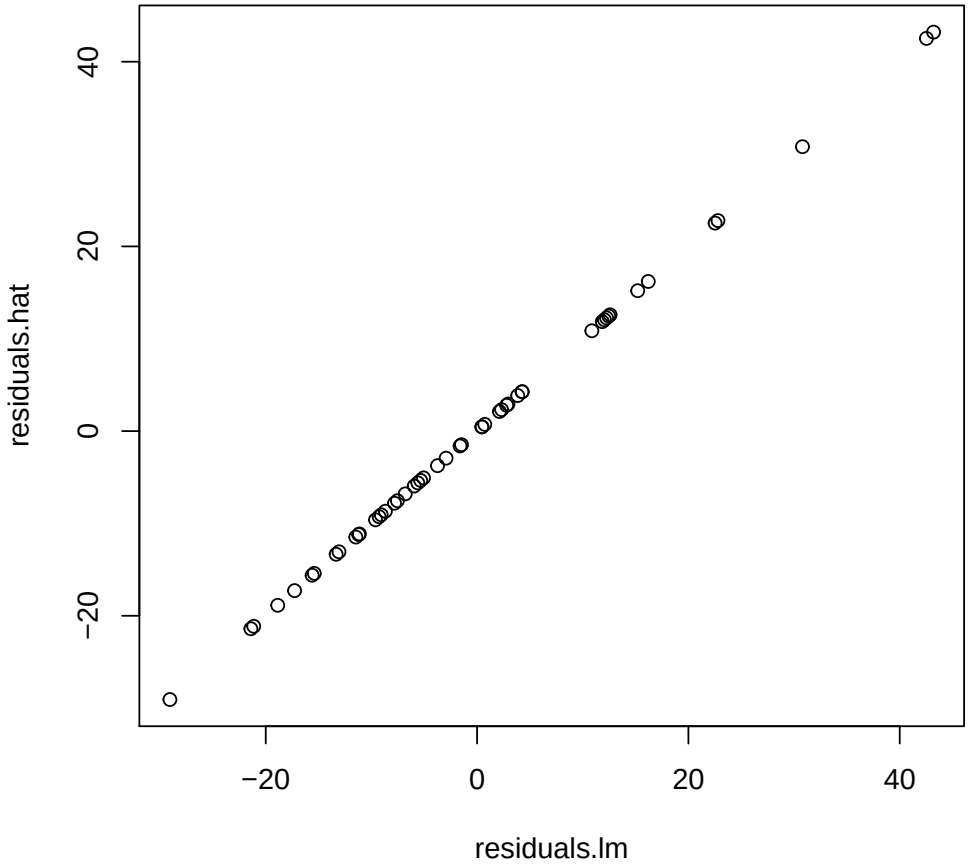
(a) $\hat{\beta}$ is given as

```
> # creating vector of 1's for intercept
> x1<-rep(1,length(speed))
> # creating design matrix
> X=matrix(0,nrow=length(speed), ncol=2)
> X[,1]<-x1
> X[,2]<-speed
> # calculating estimated parameters using design matrix and Y (distance).
> beta.hat<-(solve(t(X)%*%X))%*%(t(X))%*%distance
> beta.hat

           [,1]
[1,] -17.579095
[2,]   3.932409
```

(b) The residuals are calculated as

```
> # calculating residuals using matrix operations
> residuals.hat<-distance-(X%*%beta.hat)
> # residuals from lm
> residuals.lm<-fit$resid
> # plotting residuals.hat vs. residuals.lm
> plot(residuals.lm,residuals.hat)
```



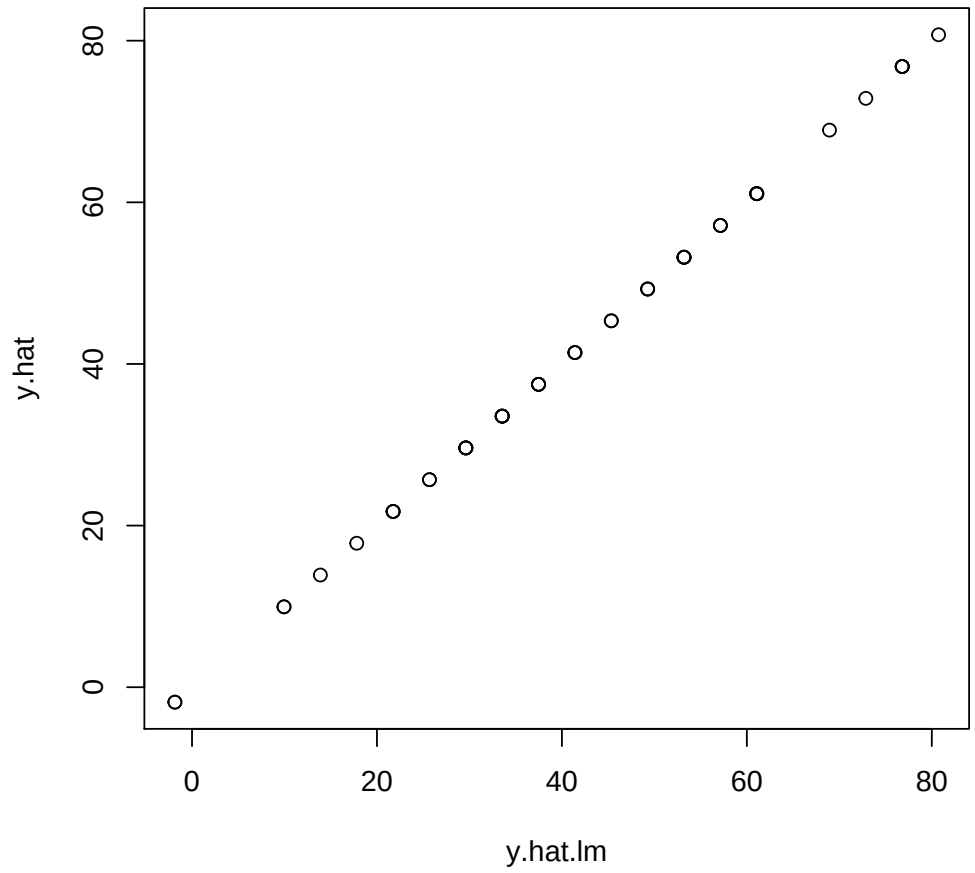
(c) $\hat{\sigma^2}$ is calculated as

```
> # calculating estimated variance
> n<-length(distance)
> p<-2
> var.hat<-(1/(n-p))*(t(residuals.hat)%*%(residuals.hat))
> var.hat
```

```
          [,1]
[1,] 236.5317
```

(d) $\hat{y}$ are calculated as

```
> # calculating Hat Matrix
> Hat.mat<-X%*%(solve(t(X)%*%X))%*%(t(X))
> # calculated estimated response
> y.hat<-Hat.mat%*%distance
> # estimated response from lm
> y.hat.lm<-fit$fitted.values
> # plotting y.hat vs. y.hat.lm
> plot(y.hat.lm,y.hat)
```



(e) se_k for the parameters are calculated as

```
> # calculating variance of estimated parameters
> sqrt.beta.var.hat<-rep(0,2)
> L<-(solve(t(X)%*%X))
> for (i in 1:2){
+   sqrt.beta.var.hat[i]<-sqrt(var.hat*L[i,i])
+ }
> sqrt.beta.var.hat

[1] 6.7584402 0.4155128
```

(f) The p-values are calculated as:

```
> # calculating p-values
> p.values<-rep(0,2)
> for (i in 1:2){
+   p.values[i]<-2*pt((-abs(beta.hat[i])/sqrt.beta.var.hat[i]), df=n-p)
+ }
> p.values

[1] 1.231882e-02 1.489836e-12
```

(g) The coefficient of determination can be calculated as:

```
> R.sq<-(t(y.hat-mean(y.hat))%*%(y.hat-mean(y.hat)))/(t(distance-mean(distance))%*%
> R.sq

          [,1]
[1,] 0.6510794
```