STAT GR5205 – Section 005 HW 4

Bo Rong br2498

Oct 22, 2016

```
#1.
#(a)
#H0 :Yi = beta0 + beta1xi +errori for i=1,...,n and Ha :Yi =f(xi)+errori for i=1,...,n errori are iid N
```

```
#(b)
filename <- "~/Downloads/copiers_full.txt"
copier_maintenance<- read.table(file=filename, header=T)
y <- copier_maintenance$minutes
x <- copier_maintenance$copiers
#Let Y be the total repair time for the ith call on which there were j copiers.
#H0: Yij = beta0 + beta1xi +errorij
#Ha: Yij = mu + errorij ,errorij are iid N(0, sigma^2).
reduced <- lm(y ~ x)
full <- lm(y ~ as.factor(x) - 1)
anova(reduced, full)
```

```
## Analysis of Variance Table
##
## Model 1: y ~ x
## Model 2: y ~ as.factor(x) - 1
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     43 3416.4
## 2     35 2797.7  8    618.72 0.9676 0.4766
```

```
#The P-value is 0.4766 and thus we accept H0.
```

```
#(c)
#The test in part (b) does not detect other departures from the normal SLR model, but the results of th
```

```
#2
#(a)
filename <- "~/Downloads/copiers_full.txt"
copier_maintenance<- read.table(file=filename, header=T)
x <- copier_maintenance$copiers
y <- copier_maintenance$minutes
fit <- lm(y ~ x, data=copier_maintenance)
alpha<-.05
g <- 2
confint(fit, level=1-alpha/g)
```

```
##                  1.25 %   98.75 %
## (Intercept) -7.092642  5.932329
## x           13.913221 16.157275
```

```r
#Hence the 95% confident of beta0 is between -7.092642 and 5.932329 and the 95% confident of beta1 is b
```

```r
#(b)
#The intervals in part (a) support this claim,because beta0=0 is in the interval (-7.092642,5.932329)
#and beta1=14 is in the interval (13.913221, 16.157275).
```

```r
#(c)
alpha <- .05
g <- 3
predict(fit,data.frame(x=c(3,5,7)),interval="confidence",level=1-alpha/g)
```

```
##         fit       lwr       upr
## 1  44.52559  40.35270  48.69848
## 2  74.59608  71.28313  77.90904
## 3 104.66658 100.65092 108.68224
```

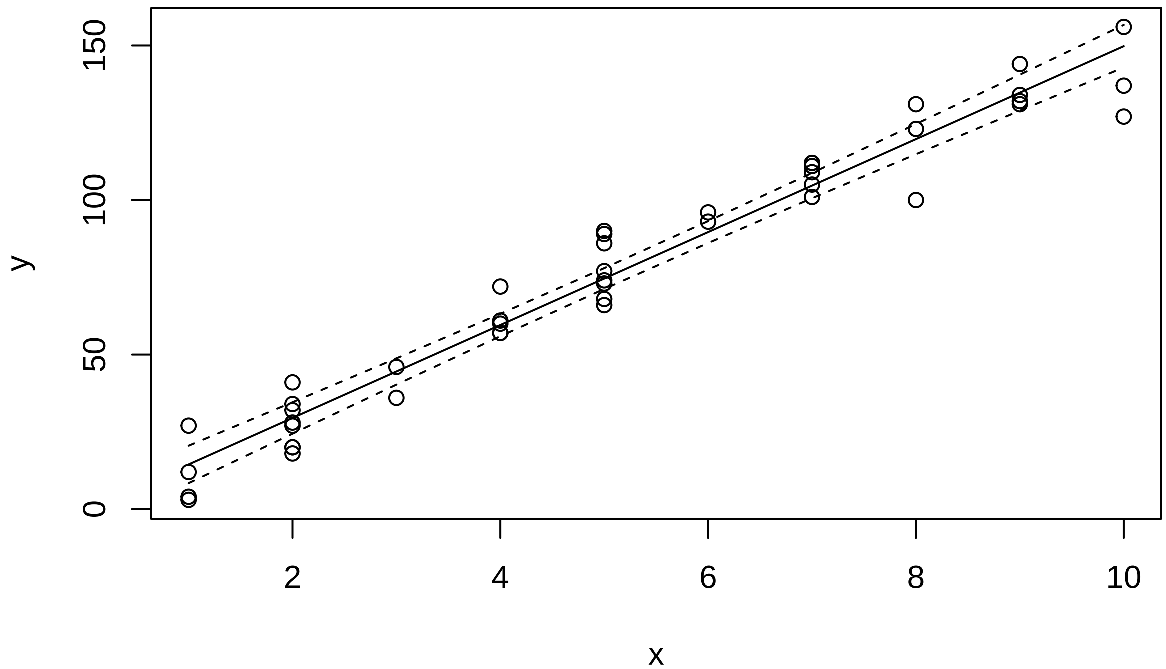
```r
#Thus the 95% confidence interval for number of minutes spent when there are 3, 5,
#and 7 copiers to be serviced are (40.35270,48.69848),(71.28313,  77.90904) and (100.65092,108.68224) r
```

```r
#(d)
alpha <- .05
g <- 2
predict(fit, data.frame(x=c(4,6)), interval="prediction", level=1-alpha/g)
```

```
##        fit      lwr       upr
## 1 59.56084 38.59230  80.52937
## 2 89.63133 68.67614 110.58652
```

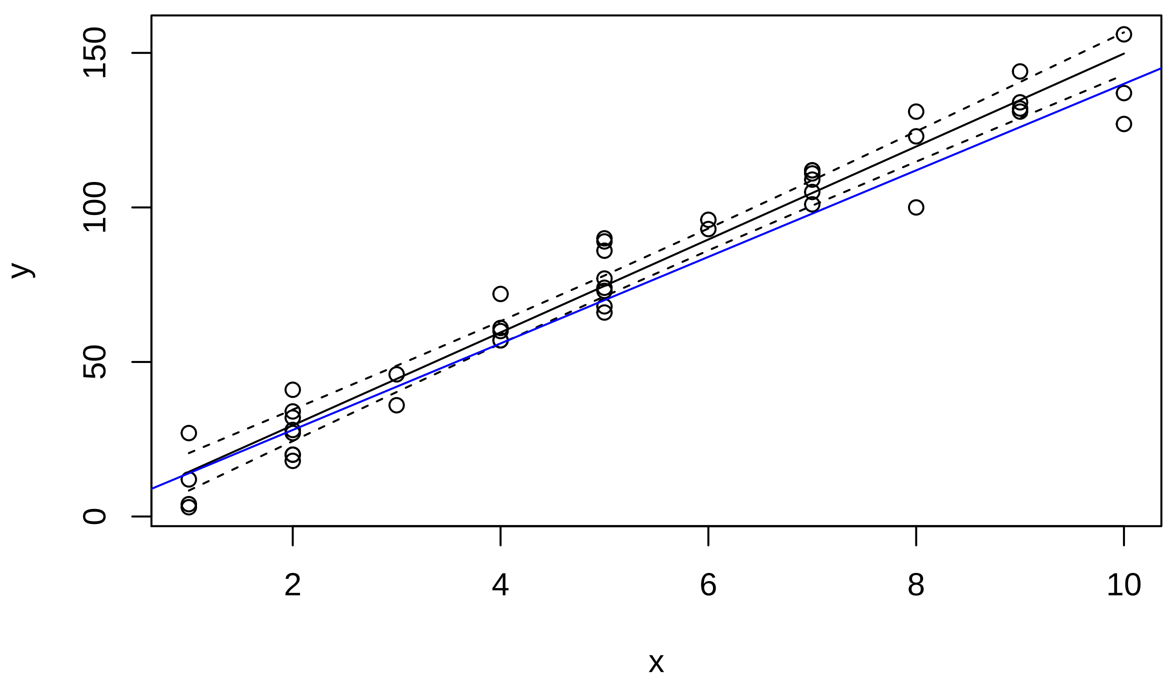
```r
#Thus the 95% confidence interval for number of minutes spent with 4 and 6 copiers to be serviced
#are (38.59230,80.52937) and (68.67614,110.58652) respectively.
```

```r
#(e)
plot(y ~ x, data=copier_maintenance)
preds <- predict(fit,data.frame(x=1:10),se.fit=T)
W.H <- sqrt(2 * qf(.95, df1=2, df2=43))
lines(1:10, preds$fit)
lines(1:10, preds$fit + W.H*preds$se.fit, lty=2.5)
lines(1:10, preds$fit - W.H*preds$se.fit, lty=2.5)
```




```
#The 95% confident interval of mean function falls within the dashed lines.
```

```
#(f)
plot(y ~ x, data=copier_maintenance)
preds <- predict(fit,data.frame(x=1:10),se.fit=T)
W.H <- sqrt(2 * qf(.95, df1=2, df2=43))
lines(1:10, preds$fit)
lines(1:10, preds$fit + W.H*preds$se.fit, lty=2.5)
lines(1:10, preds$fit - W.H*preds$se.fit, lty=2.5)
abline(0,14,col="blue")
```



```
#The line doesn't fall within the WH confidence band,thus the data are not consistent with claim.
```

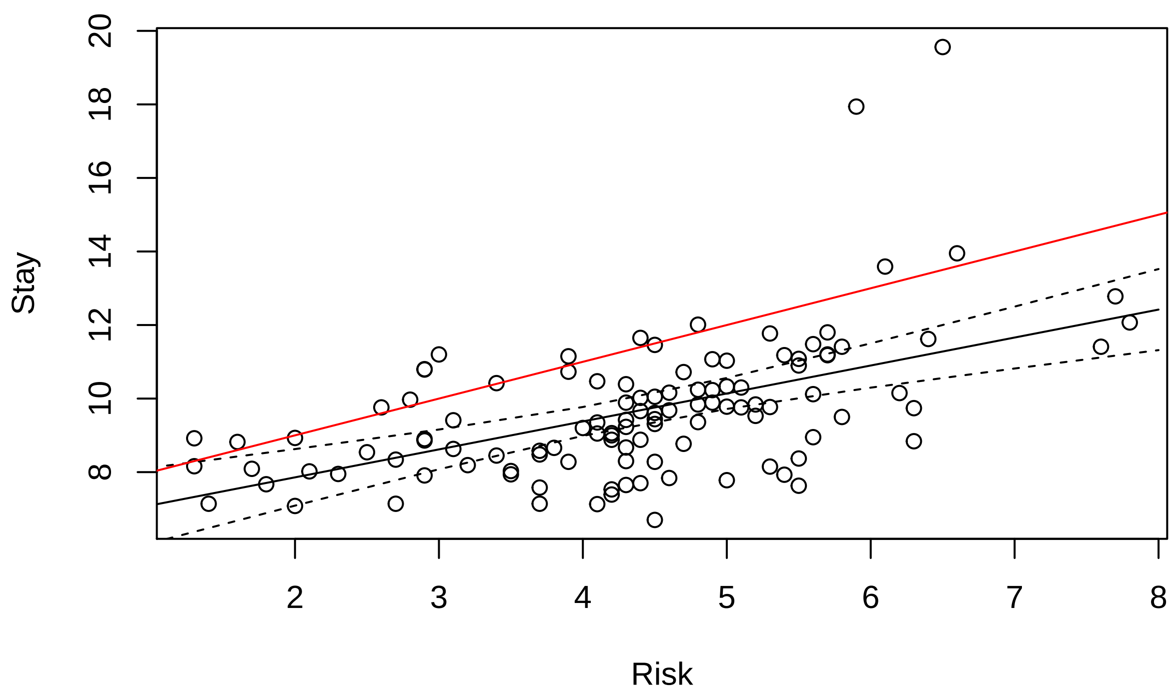
```
#3.
#(a)
filename <- "~/Downloads/SENIC.txt"
SENIC <- read.table(file=filename, header=T)
fit.Risk <- lm(Stay ~ Risk, data=SENIC)
alpha<-.05
g <- 2
confint(fit.Risk, level=1-alpha/g)
```

```
##                 1.25 %  98.75 %
## (Intercept) 5.1523372 7.521236
## Risk        0.5003816 1.020460
```

```
#Hence the 95% confident of beta1 is between 0.5003816 and 1.020460 and the 95% confident of beta0 is b
```

```
#(b)
#way 1: Since beta1=1 and beta0=7 are in the interval from part a, the data are consistent with the cla

preds <- predict(fit.Risk,data.frame(Risk=1:8), se.fit=T)
W.H <- sqrt(2 * qf(.95, df1=2, df2=111))
plot(Stay ~ Risk, data=SENIC)
lines(1:8, preds$fit)
lines(1:8, preds$fit + W.H*preds$se.fit, lty=2)
lines(1:8, preds$fit - W.H*preds$se.fit, lty=2)
abline(7,1, col="red")
```



```
#way 2: the line doesn't fall within the WH confidence band,thus the data are not consistent with claim
```

```
#(c)
alpha <- .05
g <- 4
predict(fit.Risk,data.frame(Risk=2:5),interval="confidence",level=1-alpha/g)
```

```
##         fit      lwr       upr
## 1  7.857628 7.071051  8.644206
## 2  8.618049 8.065356  9.170743
## 3  9.378470 8.977088  9.779852
## 4 10.138891 9.708059 10.569723
```

```
#Thus the 95% confident intervals of the mean stay at hospitals with infection risk x = 2,3,4 and 5
#are (7.071051,8.644206), (8.065356,9.170743), (8.977088,9.779852) and (9.708059 10.569723) respectively
```

```
#4
#(a)
filename <- "~/Downloads/copiers_full.txt"
copier_maintenance<- read.table(file=filename, header=T)
x <- copier_maintenance$copiers
y <- copier_maintenance$minutes
X<-cbind(rep(1,length(x)),x)
beta_value<-solve(t(X)%*%X)%*%t(X)%*%y
beta_value<-as.vector(beta_value)
beta_value
```

```
## [1] -0.5801567 15.0352480
```

```
#Hence the coefficients are -0.5801567 and 15.0352480.
```

```
#(b)
y_hat<-X%*%beta_value
y_hat<-as.vector(y_hat)
y_hat
```

```
##  [1]  29.49034  59.56084  44.52559  29.49034  14.45509 149.77232  74.59608
##  [8]  74.59608  14.45509  29.49034 134.73708 149.77232  89.63133  44.52559
## [15]  59.56084 119.70183 104.66658 119.70183 149.77232  59.56084  74.59608
## [22] 104.66658 104.66658  74.59608 134.73708 104.66658  29.49034  74.59608
## [29] 104.66658  89.63133 119.70183  74.59608  29.49034  29.49034  14.45509
## [36]  59.56084  74.59608 134.73708 104.66658  14.45509 134.73708  29.49034
## [43]  29.49034  59.56084  74.59608
```

```
e<-y-y_hat
e
```

```
##  [1]  -9.4903394   0.4391645   1.4744125  11.5096606  -2.4550914
##  [6] -12.7723238  -6.5960836  14.4039164 -10.4550914   2.5096606
## [11]   9.2629243   6.2276762   3.3686684  -8.5255875  12.4391645
```

```
## [16] -19.7018277   0.3334204  11.2981723 -22.7723238  -2.5608355
## [21]  -8.5960836  -3.6665796   4.3334204  -0.5960836  -0.7370757
## [26]   7.3334204 -11.4903394  -1.5960836   6.3334204   6.3686684
## [31]   3.2981723  15.4039164  -9.4903394  -1.4903394 -11.4550914
## [36]  -2.5608355  11.4039164  -2.7370757   7.3334204  12.5449086
## [41]  -3.7370757   4.5096606  -2.4903394   1.4391645   2.4039164
```

```r
#(c)
J<-matrix(rep(1, length(y)^2), length(y), length(y))
SSTO<-t(y)%*%y-1/length(y)*t(y)%*%J%*%y
SSTO<-as.numeric(SSTO)
SSTO
```

```
## [1] 80376.8
```

```r
SSE<-t(e)%*%e
SSE<-as.numeric(SSE)
SSE
```

```
## [1] 3416.377
```

```r
#(d)
MSE<-SSE/(length(y)-2)
MSE<-as.numeric(MSE)
variance_covariance_matrix_of_b<-MSE * solve(t(X) %*% X)
variance_covariance_matrix_of_b
```

```
##                     x
##    7.862086 -1.1927966
## x -1.192797  0.2333733
```

```r
#(e)
SE.fit<-sqrt(MSE*(1/length(x) + (6-mean(x))^2 / sum((x-mean(x))^2)))
t(beta_value) %*% c(1,6) + c(-1,1)*qt(.975,length(x)-2)*SE.fit
```

```
## [1] 86.81520 92.44746
```

```r
#The 95% confidence interval is (86.81520,92.44746).
```

```r
#(f)
SE.pred <- sqrt(MSE*(1+1/length(x)+(6-mean(x))^2/sum((x-mean(x))^2)))
t(beta_value)%*%c(1,6)+c(-1,1)*qt(.975,length(x)-2)*SE.pred
```

```
## [1]  71.43628 107.82639
```

```r
#The 95% prediction interval is (71.43628,107.82639).
```