HW1

Sandra Villamar and Shobhit Dronamraju

2023-01-19

```
library(ggplot2)
```

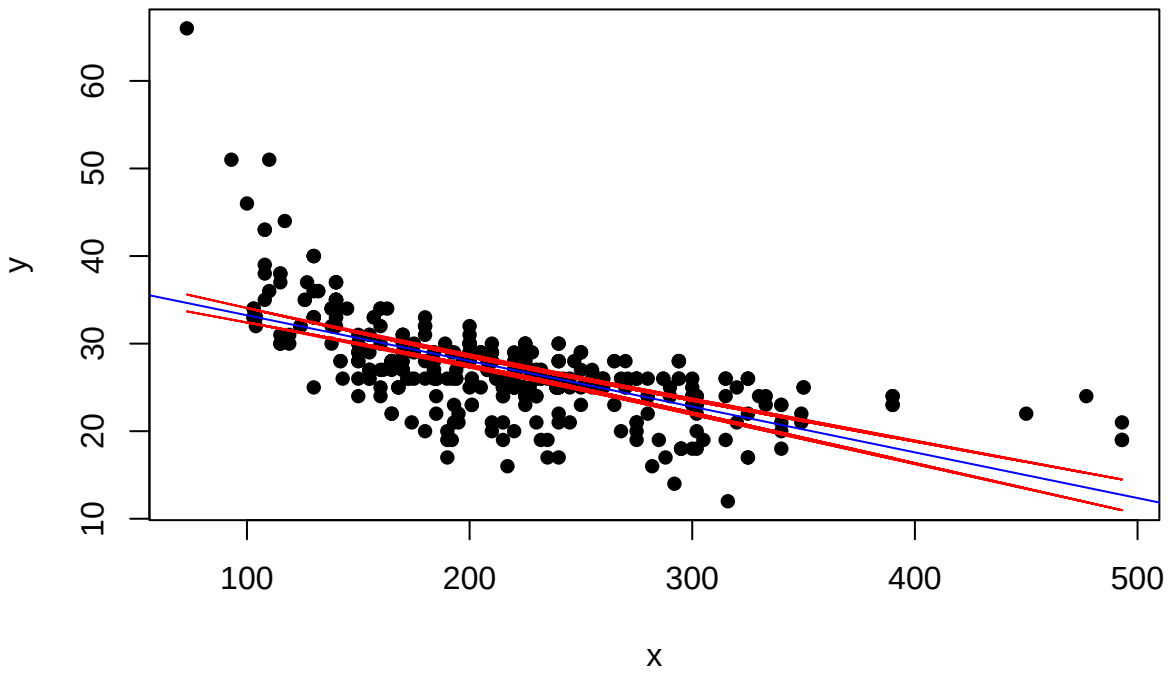
1. Write a function *confBand(x, y, conf=0.95)* taking in a predictor vector x and a response vector y and return a plot with (x, y) points, the least squares line, and the confidence band at level *conf*. Apply the function to hp and mpg from 04cars dataset.

```
confBand <- function(x, y, conf=0.95){

  # compute confidence band lines
  fit = lm(y ~ x)
  pred_interval = predict(fit, data.frame(x), interval='confidence')
  lower = pred_interval[,2]
  upper = pred_interval[,3]

  # plot
  plot(x, y, pch=16)  # points
  abline(fit$coefficients[1], fit$coefficients[2], col='blue') # reg line
  points(x, lower, col='red', type='l')  # lower band
  points(x, upper, col='red', type='l')  # upper band
}
```

```
load("HW1/04cars.rda")  # load cars dataset to "dat"
dat = dat[,c(13,15,16,18,19)]  # extract selected variables
dat = dat[complete.cases(dat),]  # extract complete cases
names(dat) = c("hp","mpg","wt","len","wd")  # abbreviate names
str(dat)
```

```
## 'data.frame':    387 obs. of  5 variables:
##  $ hp : num  103 103 140 140 140 132 132 130 110 130 ...
##  $ mpg: num  34 34 37 37 37 36 36 33 36 33 ...
##  $ wt : num  2370 2348 2617 2676 2617 ...
##  $ len: num  167 153 183 183 183 174 174 168 168 168 ...
##  $ wd : num  66 66 69 68 69 67 67 67 67 67 ...
```

```
confBand(dat$hp, dat$mpg)  # apply function
```



2. Let n = 100. Draw n i.i.d samples from *Unif(0, 1)* which stay fixed in what follows. Repeat the following experiment N=1000 times.

- Generate $y_i = 1 + x_i + \epsilon_i$ with ϵ_i i.i.d. *N(0, 0.2)*
- Compute the 99% confidence band and record whether or not it contains the true line or not

Summarize the result of this numerical experiment by returning the proportion of times (out of N) that the confidence band contained the true line.

```
n = 100
N = 1000
x = runif(n, 0, 1)
y_true = 1 + x
p = 1
count = 0

for(i in 1:N){

  # setting up formula
  e = rnorm(100, 0, sqrt(0.2))
  y = 1 + x + e

  # computing variables for confidence band
  fit = lm(y~x)
  pred = predict(fit, data.frame(x), se.fit=TRUE)
```

```
  y_hat = pred$fit
  se_hat = pred$se.fit

  # compute confidence band
  interval = sqrt((p + 1) * qf(0.99, p + 1, n - p - 1)) * se_hat
  lower = y_hat - interval
  upper = y_hat + interval

  # check if true line within band
  if(all(lower <= y_true) && all(y_true <= upper)){
    count = count + 1
  }
}

proportion = count/N
```

The experiment produced the proportion: 0.993

Contributions:

We worked on all parts of this assignment together.