# Exercise Regression 2.2 Using R properly

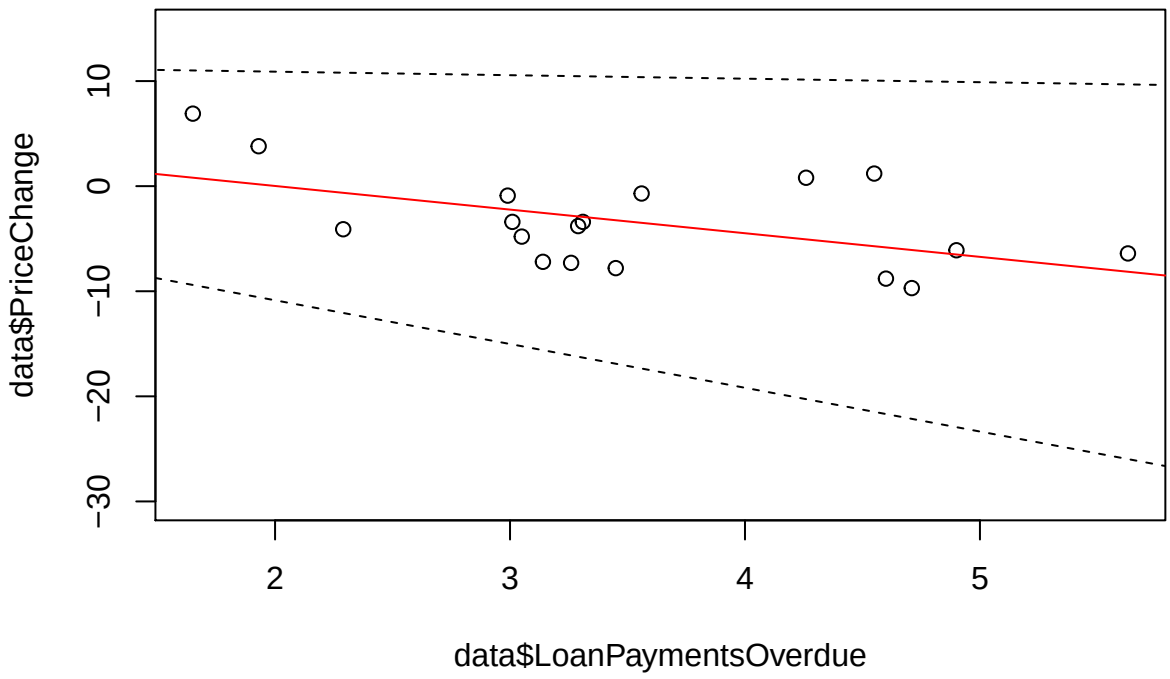
*Emil H. Andersen*

*May 20, 2016*

The exercise uses information from the data set indicator.txt

## Task a)

```
data <- read.table("indicators.txt",header=TRUE)
a <- lm(data$PriceChange~data$LoanPaymentsOverdue)
plot(data$LoanPaymentsOverdue,data$PriceChange,ylim=c(-30,15))
abline(a, col="red")
x <- seq(0,10,0.1)
ytop <- 11.5611000 -0.3335853*x
ybot <- -2.532112 -4.163454*x
lines(x,ytop,lty=2)
lines(x,ybot,lty=2)
```




We use information from the summary and confint functions of R.

```r
summary(a)
```

```
##
## Call:
## lm(formula = data$PriceChange ~ data$LoanPaymentsOverdue)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.6541 -3.3419 -0.6944  2.5288  6.9163
##
## Coefficients:
##                           Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 4.5145     3.3240   1.358   0.1933
## data$LoanPaymentsOverdue   -2.2485     0.9033  -2.489   0.0242 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.954 on 16 degrees of freedom
## Multiple R-squared:  0.2792, Adjusted R-squared:  0.2341
## F-statistic: 6.196 on 1 and 16 DF,  p-value: 0.02419
```

```r
confint(a,level=0.95)
```

```
##                               2.5 %     97.5 %
## (Intercept)              -2.532112 11.5611000
## data$LoanPaymentsOverdue -4.163454 -0.3335853
```

We see from the confint function, that the 95% confidence interval for $\beta_1$ is $[-4.163454, -0.3335853]$. We can say that we are 95% confident that there is a negative linear association, since this entire interval shows a negative decrease as LoanPaymentsOverdue increases. So this is definitely a negative linear association.

# Task b)

Using information from summary and confint, we get

```r
ymean <- 4.5145 -2.2485*4
ybotCI <- -2.532112 - 4.163454*4
ytopCI <- 11.5611000 - 0.3335853*4
c(ymean,ybotCI,ytopCI)
```

```
## [1]  -4.47950 -19.18593  10.22676
```

We observe from this estimation, that $E(Y|X = x) = -4.47950$, and the 95% confidence interval is $[-19.18593, 10.22676]$.

We can conclude that 0 is a reasonable answer for $E(Y|X = x)$ because of the large confidence interval, which happens to contain 0.