Exercise Regression 2.2

Emil H. Andersen

May 18, 2016

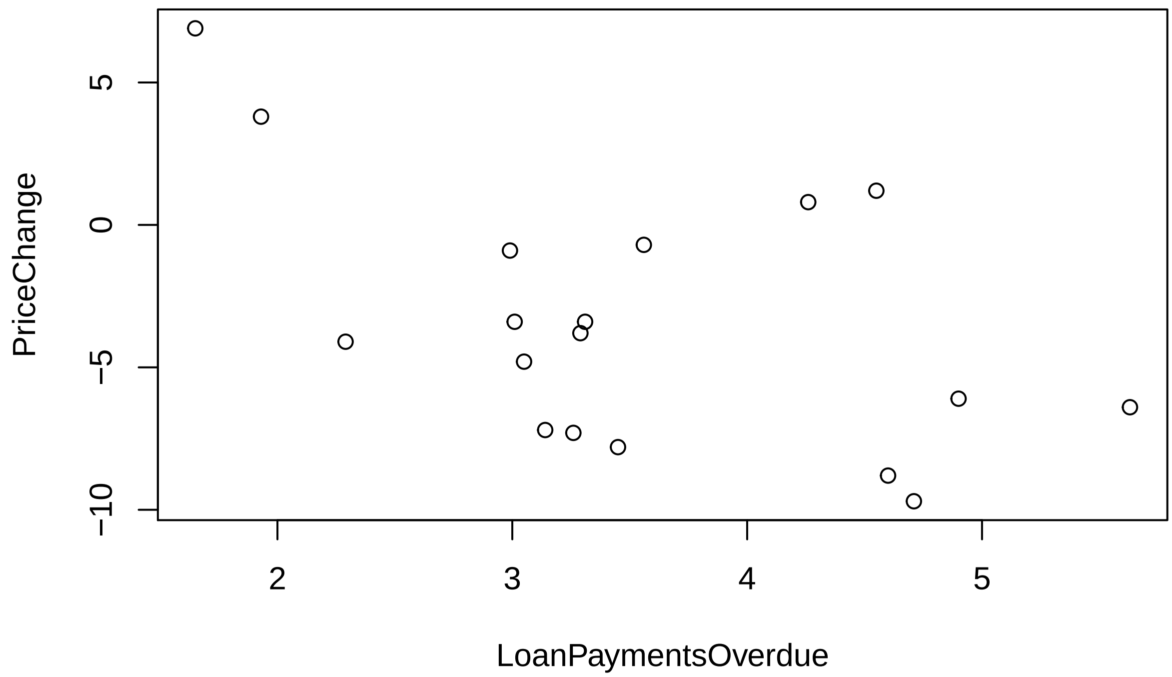
The exercise uses information from the data set indicator.txt

Task a)

```
data <- read.table("indicators.txt",header=TRUE)
data
```

```
##              MetroArea PriceChange LoanPaymentsOverdue
## 1              Atlanta         1.2                4.55
## 2               Boston        -3.4                3.31
## 3              Chicago        -0.9                2.99
## 4               Dallas         0.8                4.26
## 5               Denver        -0.7                3.56
## 6              Detroit        -9.7                4.71
## 7             LasVegas        -6.1                4.90
## 8           LosAngeles        -4.8                3.05
## 9   MiamiFt.Lauderdale        -6.4                5.63
## 10  MinneapolisStPaul         -3.4                3.01
## 11             NewYork        -3.8                3.29
## 12             Phoenix        -7.3                3.26
## 13            Portland         3.8                1.93
## 14            SanDiego        -7.8                3.45
## 15        SanFrancisco        -4.1                2.29
## 16             Seattle         6.9                1.65
## 17               Tampa        -8.8                4.60
## 18        WashingtonDC        -7.2                3.14
```

First, we look at the scatterplot of the PriceChange and LoanPaymentsOverdue You can also embed plots, for example:



We can observe from the scatterplot that it very slightly resembles a negative relation, which is good news, since this is what we want to show!

Now, we fit a linear model based on this data

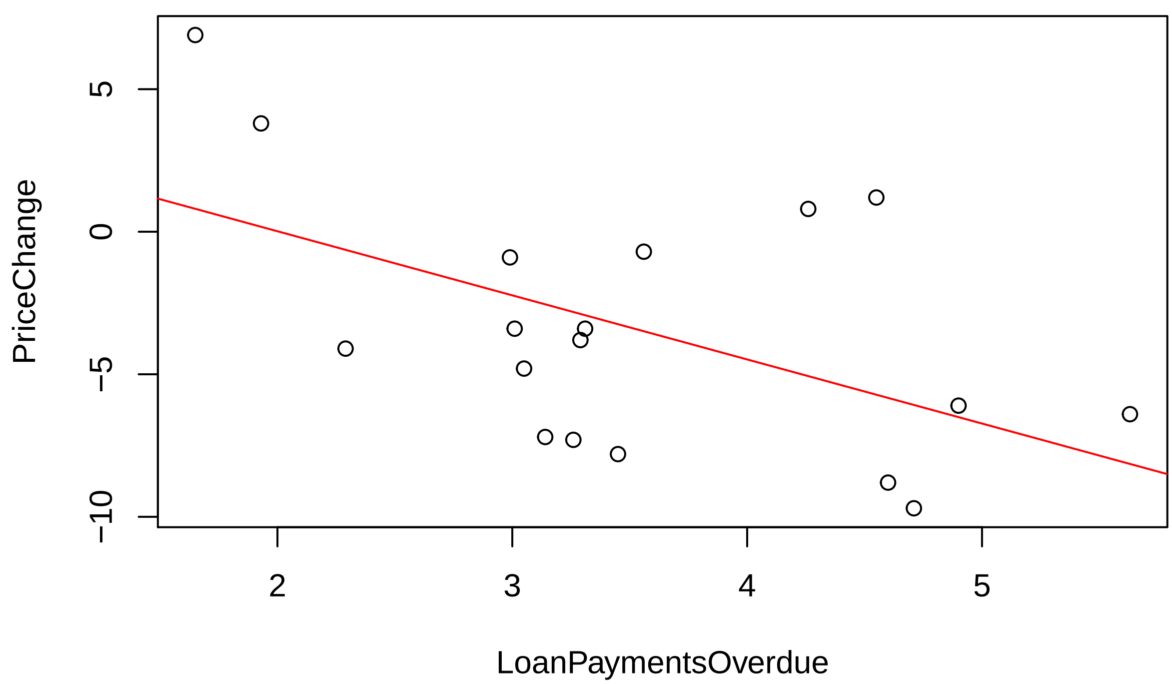
```
fit <- lm(data$PriceChange~data$LoanPaymentsOverdue)
summary(fit)
```

```
##
## 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
```

We can observe from this summary, specifically from the intercept estimate and the "data$LoanPaymentsOverdue" estimates, that the least squares best of fit is defined as: $y = 4.5145 - 2.2485 \cdot x$

```
x <- seq(0,7,0.1)
y <- 4.5145 - 2.2485*x
```

Plotting this on top of the scatter plot results in:



Now, we calculate the residuals, to use to calculate the standard error of β_1:

```
y2 <- 4.5145 - 2.2485*data$LoanPaymentsOverdue
error <- data$PriceChange - y2
S <- sqrt((1/(length(error)-2))*sum(error^2))
SXX <- sum((data$LoanPaymentsOverdue - mean(data$LoanPaymentsOverdue))^2)
seB <- S/sqrt(SXX)
B1 <- -2.2485
error
```

```
##  [1]  6.916175 -0.471965  1.308515  5.864110  2.790160 -3.624065  0.403150
##  [8] -2.456575  1.744555 -1.146515 -0.916935 -4.484390  3.625105 -4.557175
## [15] -3.465435  6.095525 -2.971400 -4.654210
```

```
seB
```

```
## [1] 0.9033113
```

We can also test the hypothesis $H_0 : \beta_1 = 0$ against $H_A : \beta_1 \neq 0$:

```
Tt <- (-2.2485-1)/seB
Tt
```

```
## [1] -3.596213
```

```
1-pt(Tt,18-2)
```

```
## [1] 0.9987908
```

To get a confidence interval of 95%, the slope of the regression line is given by: $(\hat{\beta}_1 - t(\alpha/2, n-2)se(\hat{\beta}_1), \hat{\beta}_1 + t(\alpha/2, n-2)se(\hat{\beta}_1))$

$= (-2.2485 - t(0.025, 16)0.9033113, -2.2485 + t(0.025, 16)0.9033113)$

Where t is a t-distribution. We have that:

```
c(-qt(0.975,16)*seB,qt(0.975,16)*seB)
```

```
## [1] -1.914934  1.914934
```

So a 95% confidence interval for B1 would be -2.2485 ± 1.914934. As such, we can see that there must be a negative linear relation Price Change and Loan Payments Overdue, even with the very large interval that it has.

Task b)

We estimate $E(Y|X = 4)$ using the fitted regression model from before:

```
y3 <- 4.5145 - 2.2485*4
y3
```

```
## [1] -4.4795
```

Then we find a 95% confidence interval for $E(Y|X = 4)$:

```
c(-qt(0.975,length(error-2))*S*sqrt(1+1/length(error)+
                                       (4-mean(data$LoanPaymentsOverdue))^2/SXX),
  qt(0.975,length(error-2))*S*sqrt(1+1/length(error)+(4-mean(data$LoanPaymentsOverdue))^2/SXX))
```

```
## [1] -8.58072  8.58072
```

So we have a confidence interval of $[-13.06022, 4.10122]$. Note that 0 is in fact in the interval, so 0% is a reasonable result for $E(Y|X = 4)$.