Homework5

Rohan Sadale

24 February 2016

```r
library(alr4)
```

```
## Warning: package 'alr4' was built under R version 3.2.3

## Warning: package 'car' was built under R version 3.2.3

## Warning: package 'effects' was built under R version 3.2.3
```

4.5

- Changing the base of logs would result in multiplying the coefficients by certain factor, but the results(same R-squared value, significance of coefficients, residual standard error) remains the same.
- So for the second part of the question, we need to multiply log(10) to the coefficient of model with log10(Y) as response.
- For example

```r
x<-c(1:10)

#Model with natural log in response
m1 <- (lm(log(x) ~ x))
summary(m1)
```

```
##
## Call:
## lm(formula = log(x) ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.47361 -0.09245  0.04142  0.16563  0.22146
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.24320    0.16313   1.491    0.174
## x            0.23041    0.02629   8.764 2.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2388 on 8 degrees of freedom
## Multiple R-squared:  0.9057, Adjusted R-squared:  0.8939
## F-statistic:  76.8 on 1 and 8 DF,  p-value: 2.253e-05
```

```r
#Model with log10 in response
m2 <- (lm(log10(x) ~ x))
summary(m2)
```

```
##
## Call:
## lm(formula = log10(x) ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.20569 -0.04015  0.01799  0.07193  0.09618
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.10562    0.07085   1.491    0.174
## x            0.10006    0.01142   8.764 2.25e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1037 on 8 degrees of freedom
## Multiple R-squared:  0.9057, Adjusted R-squared:  0.8939
## F-statistic:  76.8 on 1 and 8 DF,  p-value: 2.253e-05
```

From the coefficients we can see that if we divide coefficients of model m1 by log(10), we get same coefficients as in model m2.

4.6 As the response variable is in log, usual interpretation of effect of coefficients of linear model is not useful i.e. 1 unit change in pctUrban gives the rate of change in log (fertility).

Thus, to the change in fertility for unit change in pctUrban is -1%.

```r
100*(exp(-0.01) -1)
```

```
## [1] -0.9950166
```

4.7 The 25% increase in ppgdp will change fertility by -

```r
m1 <- lm(log(fertility) ~ log(ppgdp) + lifeExpF, UN11)
coefficients(m1)
```

```
## (Intercept)  log(ppgdp)    lifeExpF
##  3.50736168 -0.06543735 -0.02823612
```

```r
100*(exp(log(1.25) * (-0.06544)) -1)
```

```
## [1] -1.449641
```

4.9

4.9.1

- The intercept is 24697. Thus $24697 would be expected salary of male faculty. For female faculty, the expected salary would be 24697 - 3340 = $21357.

4.9.2

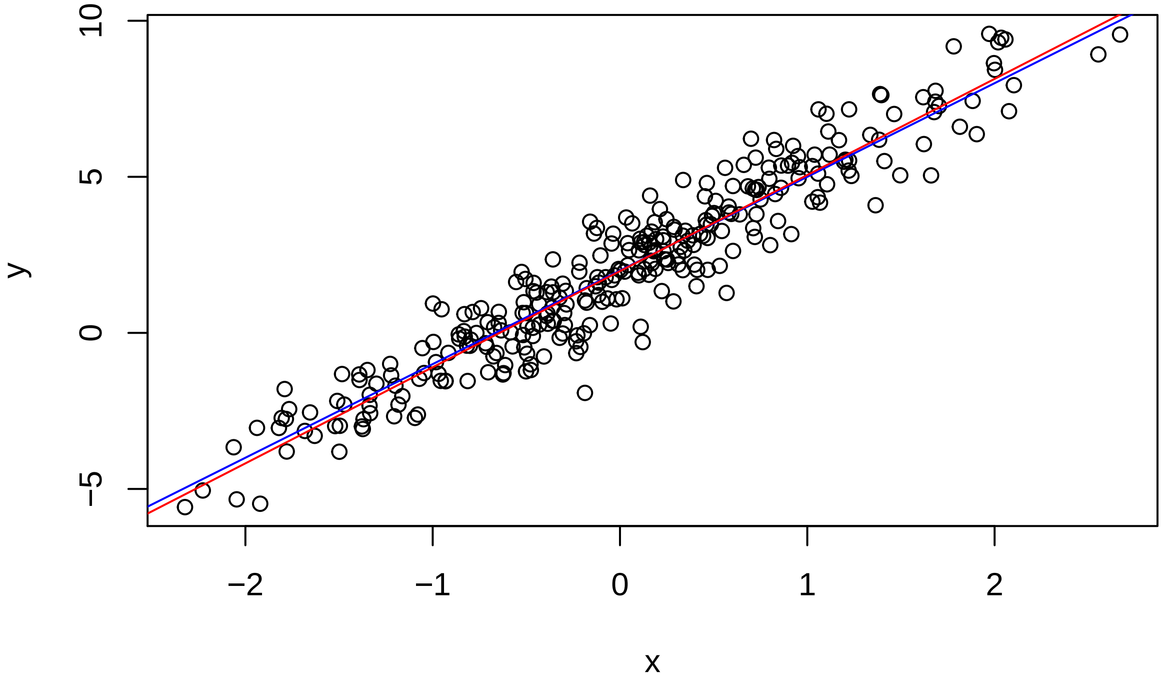
- E(Salary|Sex) = 24697 - 3340Sex E(Salary|Sex,Years) = 18065 + 201Sex + 759 Years
- Subtracting above two equations and equating, 24697 - 3340Sex = 18065 + 201Sex + 759Years Years = (24697-18065)/759 - (3340 - 201) * Sex/759 Years = 8.7 - 4.7Sex
- From this we can interpret that the data to be consistent, the average male has about 8.7 years of experience and average female has 4 years of experience.

4.12

```r
set.seed(1000)
x <- rnorm(300)
e <- rnorm(300)
y <- 2 + 3*x + e
plot(y~x)

y1 <- 2 + 3*x
m1 <- lm(y1~x)
abline(m1, col = "blue")

m2 <- lm(y ~ x)
abline(m2, col="red")
```



4.12.1

From the plot we can see that the scatter of points is approximately elliptical, and the regression line is similar to, but not exactly the same as, the major axis of the ellipse.

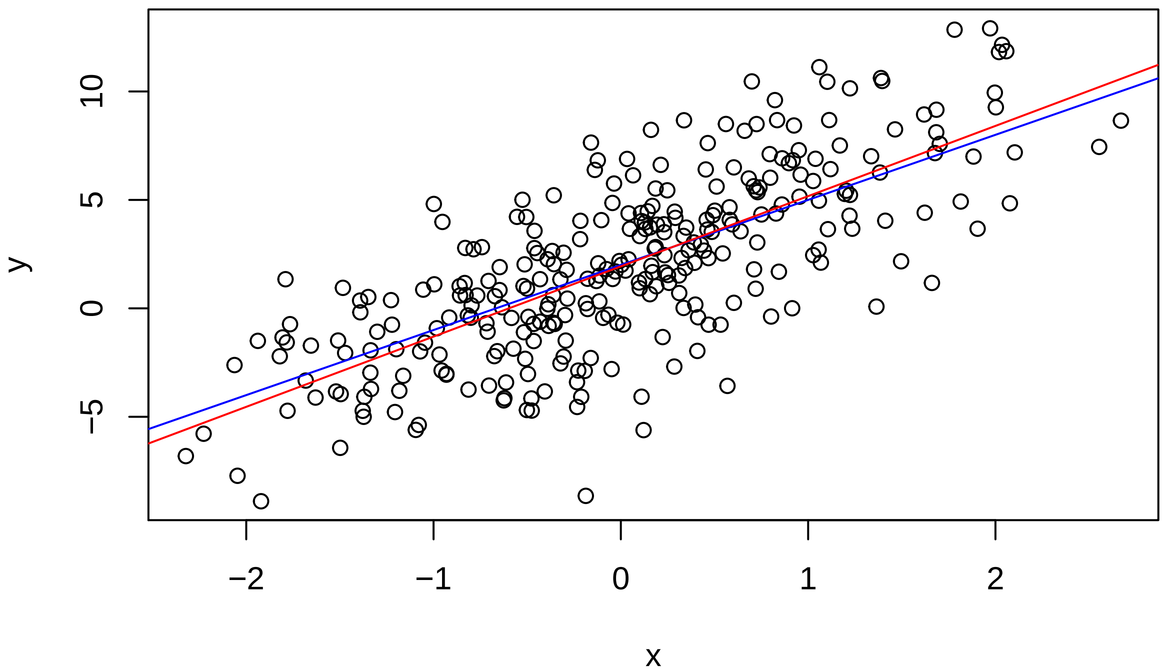
4.12.2

- For sigma = 3

```
set.seed(1000)
x <- rnorm(300)
e <- rnorm(300)
y <- 2 + 3*x + 3*e
plot(y~x)

y1 <- 2 + 3*x
m1 <- lm(y1~x)
abline(m1, col = "blue")

m2 <- lm(y ~ x)
abline(m2, col="red")
```

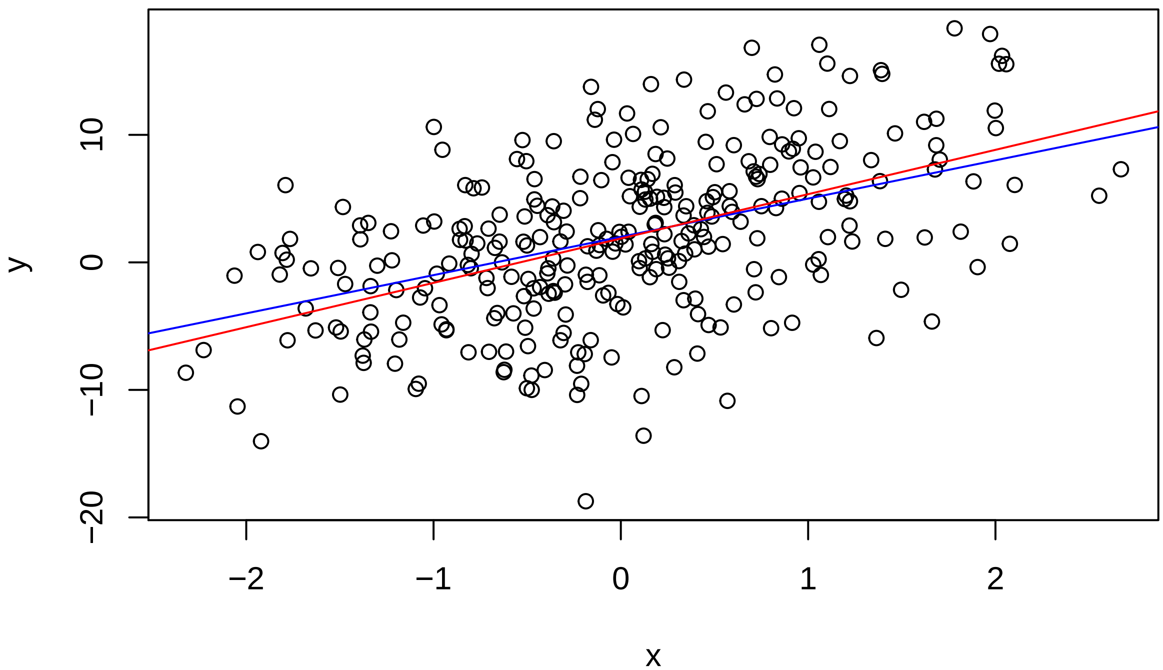


+ For sigma = 6

```
set.seed(1000)
x <- rnorm(300)
e <- rnorm(300)
y <- 2 + 3*x + 6 * e
plot(y~x)

y1 <- 2 + 3*x
m1 <- lm(y1~x)
abline(m1, col = "blue")

m2 <- lm(y ~ x)
abline(m2, col="red")
```



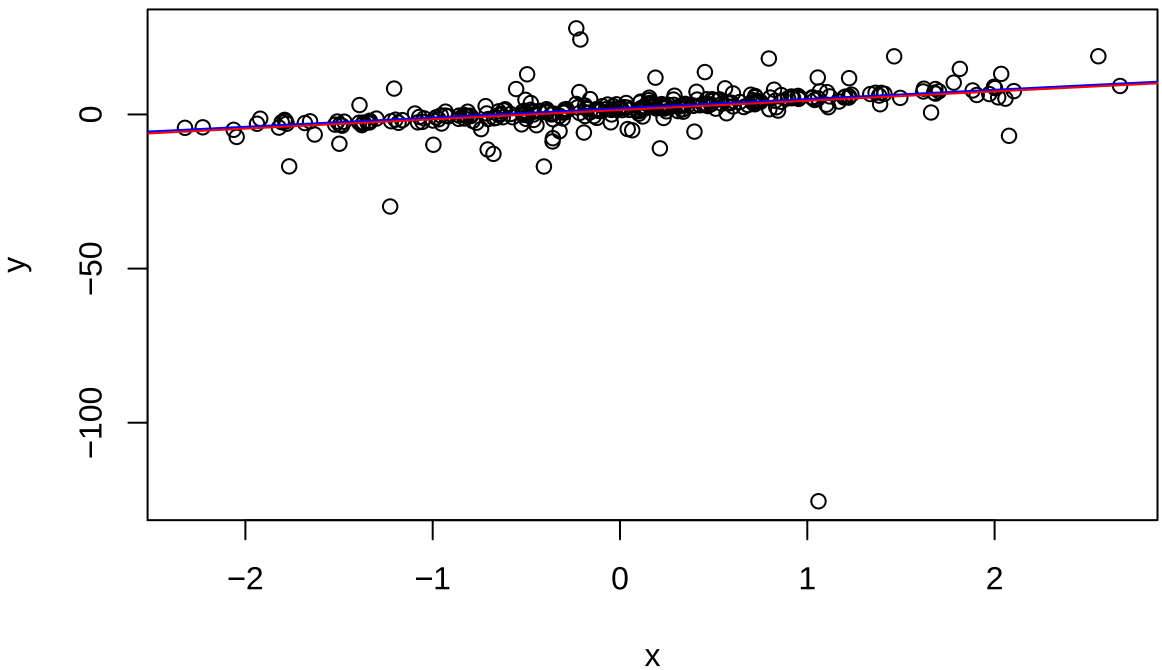
We can see from the two plots that there is an increase in spread of points along y axis as sigma increases from 3 to 6.

```
set.seed(1000)
x <- rnorm(300)
v1 <- rnorm(300)
v2 <- rnorm(300)
e <- v1/v2

y <- 2 + 3*x + e
plot(y~x)

y1 <- 2 + 3*x
m1 <- lm(y1~x)
abline(m1, col = "blue")

m2 <- lm(y ~ x)
abline(m2, col="red")
```



4.12.3