HW3

DANDONG TU

2017/9/13

1.

writing

$$c_i = \frac{(x_i - \bar{x})}{SXX}$$

then

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{SXX} = \sum(\frac{x_i - \bar{x}}{SXX})y_i = \sum c_i y_i$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = \sum(\frac{1}{n} - c_i\bar{x})y_i = \sum d_i y_i$$

with $d_i = (\frac{1}{n} - c_i x_i)$. A fitted value $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ is equal to $\sum(d_i + c_i x_I)y_i$, also a linear combination of the y_i where i=1,2...n.

2.

2.2.1

The solid line y=x means when increse in x (which is 2003 Rice price) that increse same value in y(2009 Rice price),therefore, the points above the line refer to the prices in 2009 higher than prices in 2003 relatively which may indicates that these cities are recovered from the major recession around 2006 in terms of rice price index. The points below the line refer to the prices in 2009 lower than prices in 2003 relatively which may indicates that these cities are not recoverd comparing 2009 to 2003 in terms of rice price index.

2.2.2

From the figure and considering the slope of each points, Vilnius has the largest increse in rice price since it has largest slope, and meanwhile, the city of Mumbai has the largest decrese in rice price due to the lowest slope.

2.2.3

$\hat{\beta}_1 < 1$ and from the figure it is easy to concludes that $\hat{\beta}_0 > 0$. The intersection point between ols and y=x are also presentted in the firgue. The rice prices are higher in 2009 than in 2003 before the intersection point and the prices are lower in 2009 than in 2003 after the intersection point.

2.2.4

1.Even though using minutes of labor corrects part for currency fluctuations, prevailing wage rates and local prices, there are still lots of factors should be taken account such as rice demand-supply situation, import and export policy and currency exchange rate. So that we may establish a multiple regression model to better illustrate the relations. 2.From statisticl point of view, evaluating these much information is not much simple

as the fitting simple linear regression model,however, using log-scale might be preferable to illustrate the relations

3.

Simulation

```
beta_0=10
beta_1=-2.5
n=30
```
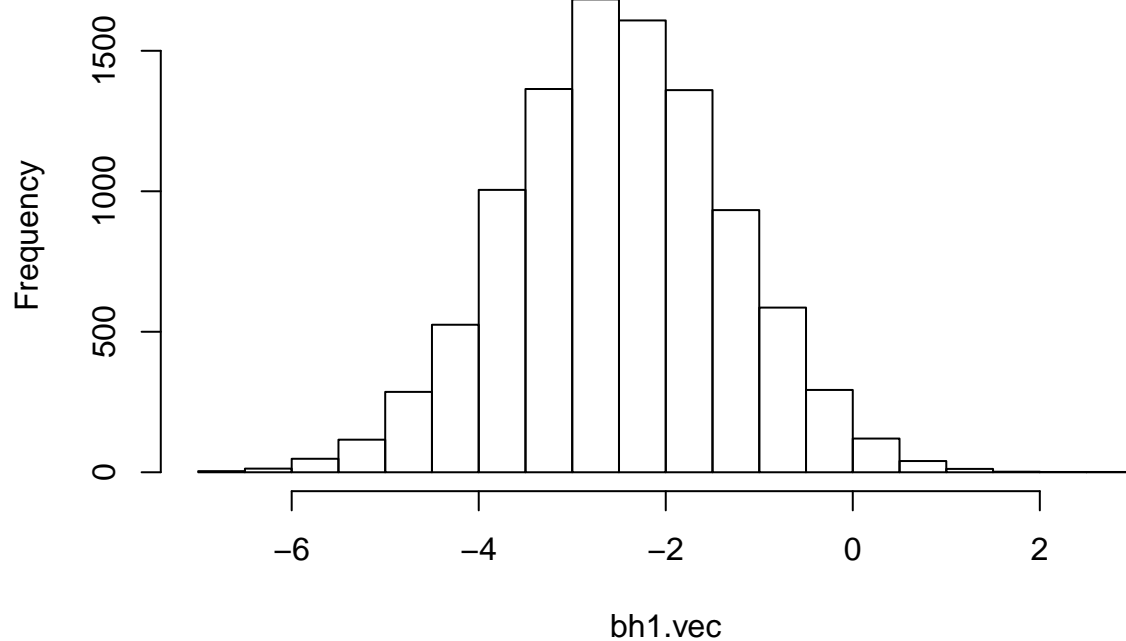
a)

```
sigma=100
i=1:n
x_i=i
```

b)

```
set.seed(1016)
x=sample(100:150,n,replace = TRUE)


e = rnorm(n, mean = 0, sd = sigma)
y = beta_0+beta_1*x+e

sim = 10000
bh1.vec = rep(0,sim)
bh0.vec = rep(0,sim)
sigma2.hat=rep(0,sim)

for (i in 1:sim){
  e = rnorm(n, mean = 0, sd = sigma)
  y = beta_0+beta_1*x+e
  mi = lm(y ~ x)
  bh1.vec[i] = coef(mi)[2]
  bh0.vec[i] = coef(mi)[1]
  RSS=sum((mi$residuals)^2)
  sigma2.hat[i]=RSS/(n-2)}
```
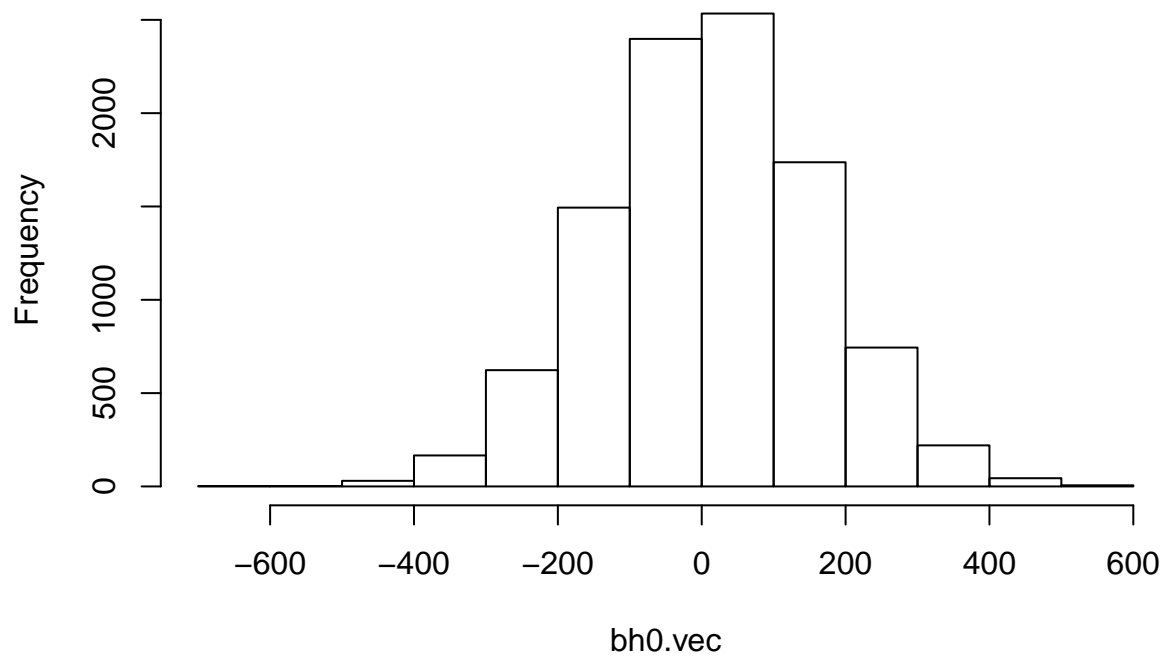
c)

```
hist(bh1.vec)
```
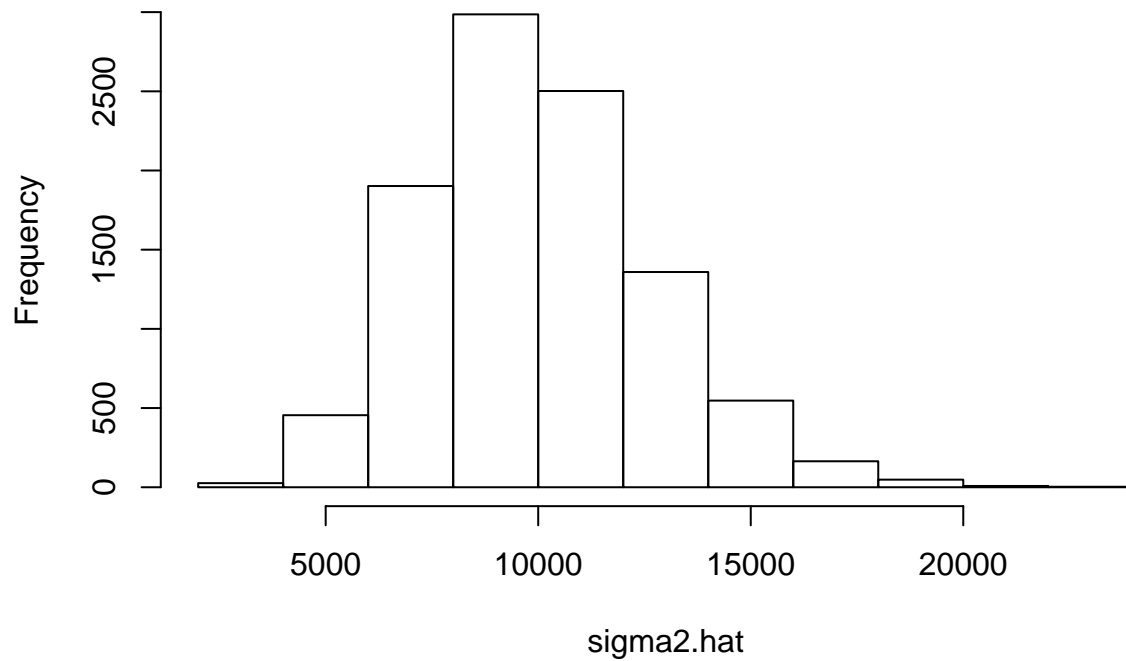
Histogram of bh1.vec


```r
hist(bh0.vec)
```

Histogram of bh0.vec


```r
hist(sigma2.hat)
```

Histogram of sigma2.hat

the shape for figures 1 and 2 based on the data from linear regression parameter estimates: betahat0, betahat1, and they are roughly bell shaped with mean of 10, -2. And the third figure sigmasquare is kind of right skewed.

d)

```
a=mean(bh1.vec)
b=mean(bh0.vec)
c=mean(sigma2.hat)

a
```

```
## [1] -2.498593
```

```
b
```

```
## [1] 9.902367
```

```
c
```

```
## [1] 9989.369
```

Compared with the true parameters. the value of averages of betahat0, betahat1 and sigmasquare is pretty much similar as the true parameters which are -2.5 10 and 100 relatively.

e)

```
(var0=var(bh0.vec))
```

```
## [1] 22616.69
```

```r
(var1=var(bh1.vec))
```

```
## [1] 1.408486
```

```r
SXX=sum((x-mean(x))^2)


Var.true=(sigma^2)/SXX

Var.true
```

```
## [1] 1.426235
```

```r
Var.true.1=sigma^2*((1/n)+(mean(x)^2)/SXX)

Var.true.1
```

```
## [1] 22928.35
```

The value for the betahat0 and betahat1 are pretty much similar with the true variances

f)

```r
i=1:n
x_i2=100*i


x1=sample(100:150,n,replace = TRUE)


e1 = rnorm(n, mean = 0, sd = sigma)
y1 = beta_0+beta_1*x1+e

sim1 = 10000
bh1.vec1 = rep(0,sim)
bh0.vec1 = rep(0,sim)
sigma2.hat1=rep(0,sim)

for (i in 1:sim){
  e1 = rnorm(n, mean = 0, sd = sigma)
  y1 = beta_0+beta_1*x+e
  mi1 = lm(y ~ x)
  bh1.vec1[i] = coef(mi1)[2]
  bh0.vec1[i] = coef(mi1)[1]
  RSS1=sum((mi1$residuals)^2)
  sigma2.hat1[i]=RSS/(n-2)}

a1=mean(bh1.vec1)
b1=mean(bh0.vec1)
c1=mean(sigma2.hat1)

a1
```

```
## [1] -3.230356
```

```
b1
```

```
## [1] 117.3689
```

```
c1
```

```
## [1] 7226.36
```

```
(var01=var(bh0.vec1))
```

```
## [1] 0
```

```
(var11=var(bh1.vec1))
```

```
## [1] 0
```

```
SXX1=sum((x1-mean(x1))^2)


Var.true1=(sigma^2)/SXX1

Var.true1
```

```
## [1] 1.923817
```

```
Var.true.11=sigma^2*((1/n)+(mean(x1)^2)/SXX1)

Var.true.11
```

```
## [1] 28967.42
```

The new sample variance of betahat0 and betaha1 are pretty close with the previous results. The reason for that is that even though the x_i became the 100times, the final value increasd 100 times, however, the sample variance stay closed.