JAE-HO YOON

THE SIMPLE LINEAR REGRESSION MODEL

LOAD DATA

```
food = read.table(url("http://www.principlesofeconometrics.com/poe4/data/dat/food.dat"),header=F)

head(food)
```

```
##       V1    V2
## 1 115.22  3.69
## 2 135.98  4.39
## 3 119.34  4.75
## 4 114.96  6.03
## 5 187.05 12.47
## 6 243.92 12.98
```

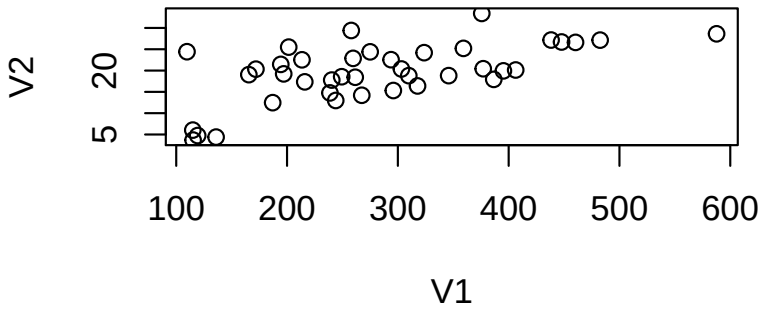
DATA Summary

```
summary(food)       # Summary statistics
```

```
##        V1              V2
##  Min.   :109.7   Min.   : 3.69
##  1st Qu.:200.4   1st Qu.:17.11
##  Median :264.5   Median :20.03
##  Mean   :283.6   Mean   :19.60
##  3rd Qu.:363.3   3rd Qu.:24.40
##  Max.   :587.7   Max.   :33.40
```

Plot Data

```
plot(food)          # Plot matrix
```



Model

$$Y_i = \alpha + \beta X_i + e_i$$

Variables

```
x = food[,2]   # Income
y = food[,1]   # Food Expenditure
```

Estimation

```
reg <- lm(y ~ x)
```

Results

```
summary(reg)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -223.025  -50.816   -6.324   67.879  212.044
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   83.416     43.410   1.922   0.0622 .
## x             10.210      2.093   4.877 1.95e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 89.52 on 38 degrees of freedom
## Multiple R-squared:  0.385,  Adjusted R-squared:  0.3688
## F-statistic: 23.79 on 1 and 38 DF,  p-value: 1.946e-05
```

b1, b2

```
b1 <- coef(reg)[[1]]
b2 <- coef(reg)[[2]]
```

```
b1
```

```
## [1] 83.416
```

```
b2
```

```
## [1] 10.20964
```

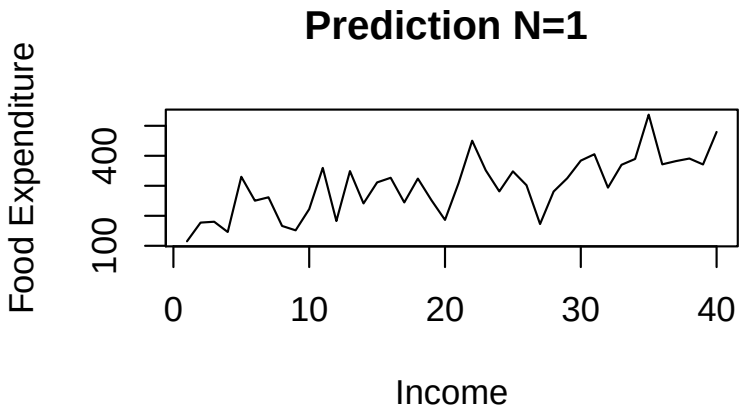
Least sqaures prediction (one time)

```
N <- 40
sde <- 89.52
y1 <- b1+b2*x+rnorm(N, mean=0, sd=sde)

y2 <- data.frame()
y2 <- cbind(y1, y)
```

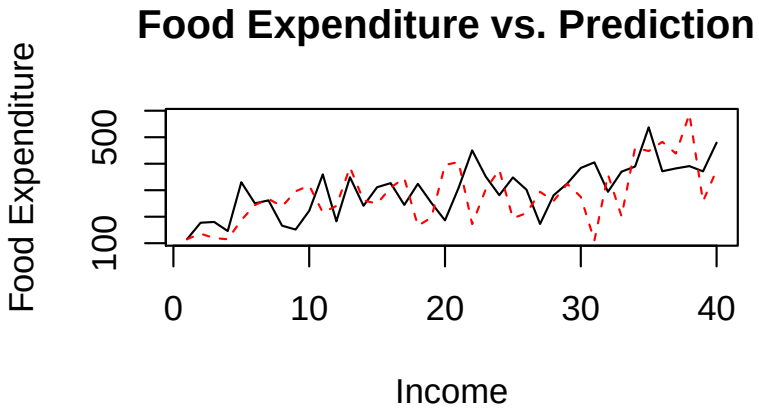
Least sqaures prediction (one time)

```
matplot(y1, type='l', col=1:40,
        xlab='Income', ylab='Food Expenditure',
        main ='Prediction N=1 ')
```



Least sqaures prediction (one time)

```
matplot(y2, type='l', col=1:40,
        xlab='Income', ylab='Food Expenditure',
        main ='Food Expenditure vs. Prediction ')
```



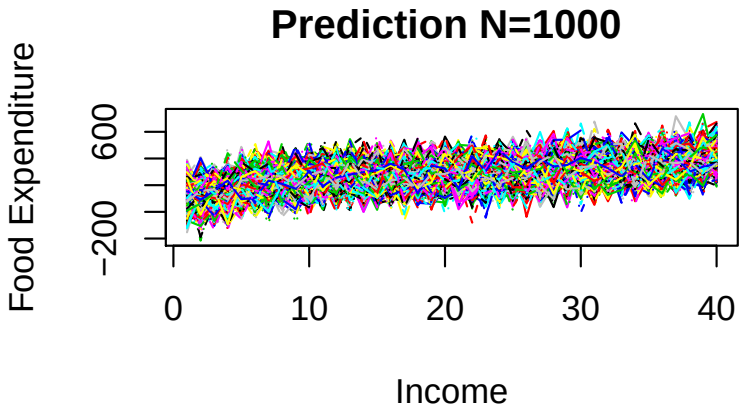
Least sqaures prediction (1,000 times)

```
b1 <- coef(reg)[[1]]
b2 <- coef(reg)[[2]]
yy <- data.frame()

trial <- 1
trials <- 1000
while(trial <= trials) {
  y3 <- b1+b2*x+rnorm(N, mean=0, sd=sde)
  yy <- rbind(yy, t(y3))
  trial <- trial + 1
}
```

Least sqaures prediction (1,000 times)

```
matplot(t(yy), type='l', col=1:40,
        xlab='Income', ylab='Food Expenditure',
        main ='Prediction N=1000 ')
```



Save DATA

sink('ch4.out')

```
# Least sqaures prediction (one time)
```

```
y1
```

```
##  [1] 114.9626 177.2920 180.2351 145.9258 330.0973 250.3515 261.6975 166.1512
##  [9] 151.5264 222.2123 359.6088 182.5448 348.9705 241.2174 311.2865 326.7293
## [17] 244.5144 324.0969 252.1103 186.2666 308.9445 450.2580 351.6664 281.2010
## [25] 348.1514 301.9825 172.5609 280.9466 325.1279 383.9049 405.1841 293.9387
## [33] 369.9601 389.4640 537.1353 371.5209 382.2521 390.9060 371.1812 479.4355
```

sink()

References

- Principles of Econometrics with R `https://bookdown.org/ccolonescu/RPoE4/`
- Principles of Econometrics `http://www.principlesofeconometrics.com/poe4/poe4.htm`
- Beowulfkorea `https://sites.google.com/site/beowulfkorea/yoon/r`