GDP2

Daning Bi

March 3, 2016

Part a

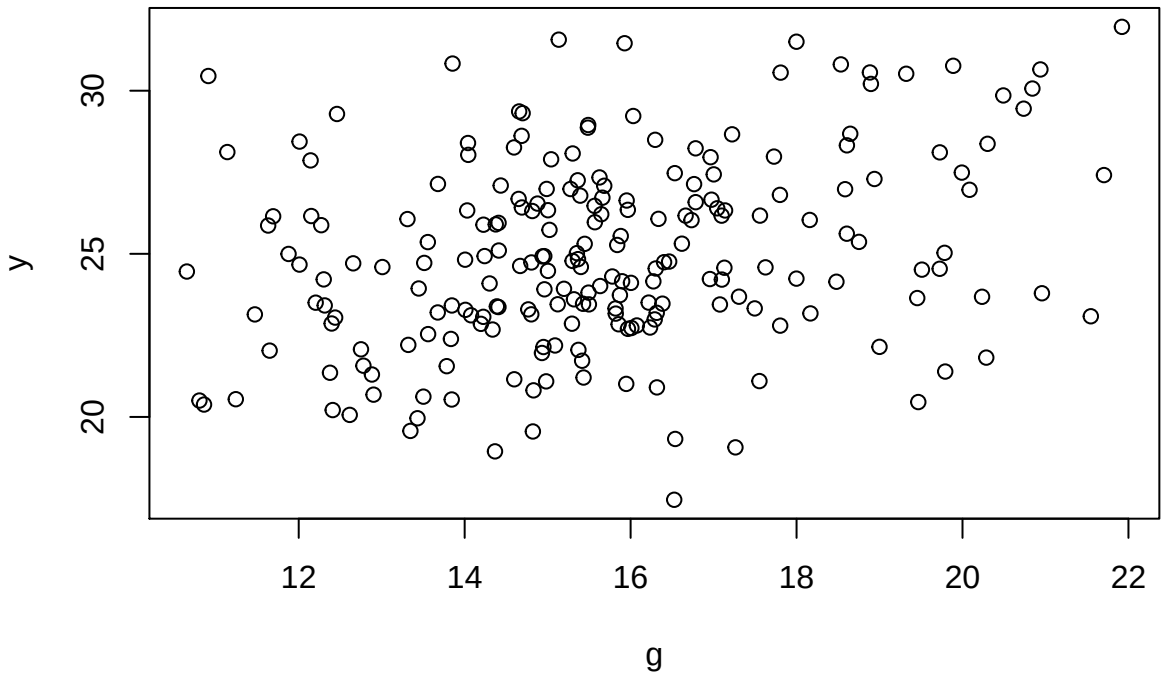
First of all we need to read the data into R:

```
setwd("D:\\ANU\\2016S1\\STAT8027\\stat8027")
gdp <- read.csv("gdp2013.csv")
labor <- read.csv("labor2013.csv")
```

Before we run the linear regression model, we need to have a check of the data. We take logritham to remove the skewness of the data, then plot the labor with gdp. The missing value has also been removed.

```
y<-as.matrix(na.omit(log(gdp[,3])))
g<-as.matrix(na.omit(log(labor[,3])))

plot(g,y)
```



According to the matrix form solution of linear regression, $

$$\beta = (\boldsymbol{X'X})^{-1}\boldsymbol{X'Y}$$

.

Hence we can solve the linear regression parameters using matrix calculation:

```
i <- as.matrix(rep(1,length(g)))
x <- cbind(i,g)
beta <- solve(t(x)%*%x)%*%t(x)%*%y
beta
```

```
##            [,1]
## [1,] 19.4467534
## [2,]  0.3531498
```

Hence we can solve that $\beta_0 = 19.447$, and $\beta_1 = 0.353$.

Then we can use `lm()` function in R to check if the answer we calculated is correct or not:

```
lr <- lm(y~g)
summary(lr)
```

```
##
## Call:
## lm(formula = y ~ g)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.8212 -1.8724 -0.0526  1.8851  7.1506
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 19.44675    1.24648  15.601  < 2e-16 ***
## g            0.35315    0.07868   4.488 1.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.769 on 216 degrees of freedom
## Multiple R-squared:  0.0853, Adjusted R-squared:  0.08107
## F-statistic: 20.14 on 1 and 216 DF,  p-value: 1.168e-05
```

```
lr$coefficients
```

```
## (Intercept)           g
##  19.4467534   0.3531498
```

As they are exactly the same, we have double checked our solution.