GDP

Daning Bi

Febuary 21, 2016

Part a

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

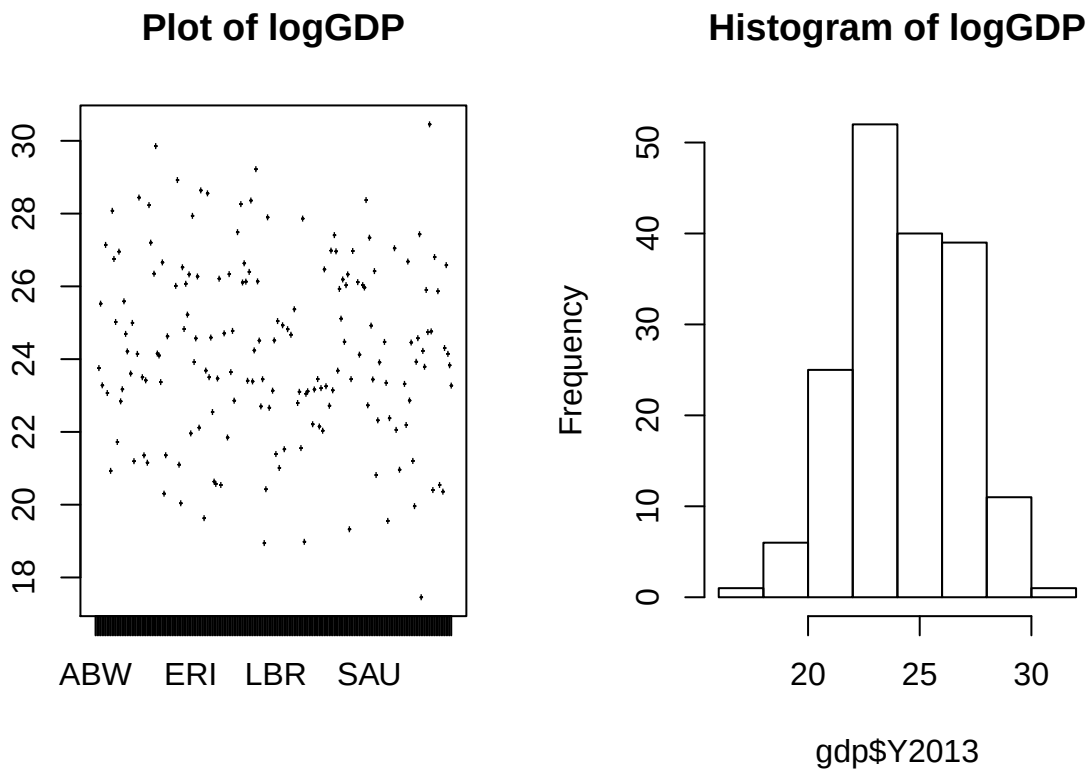
```
setwd("D:\\ANU\\2016S1\\STAT8027\\stat8027\\")
gdp <- read.table("GDP.txt", header = T)
```

Then before we plot the data, we should take natural log of GDP fisrt as the scale of data is quite big which may affect the plot:

```
gdp$Y2013<-log(gdp$Y2013)
```

After transformation, we can take a plot as well as a histogram of the log gdp:

```
par(mfrow=c(1,2))
plot(gdp$Country.Code,gdp$Y2013, main = "Plot of logGDP")
hist(gdp$Y2013, main = "Histogram of logGDP")
```



From the picture we may found that the log GDP of different countries are close to a normal distribution and the plot shows that the log GDPs are almost randomly allocated around 25.

Part b

Next we can compute a six number summary of log GDP which includes mean, median, maximum, minimum, first quantile and third quantile.

```
summary(gdp$Y2013)
```

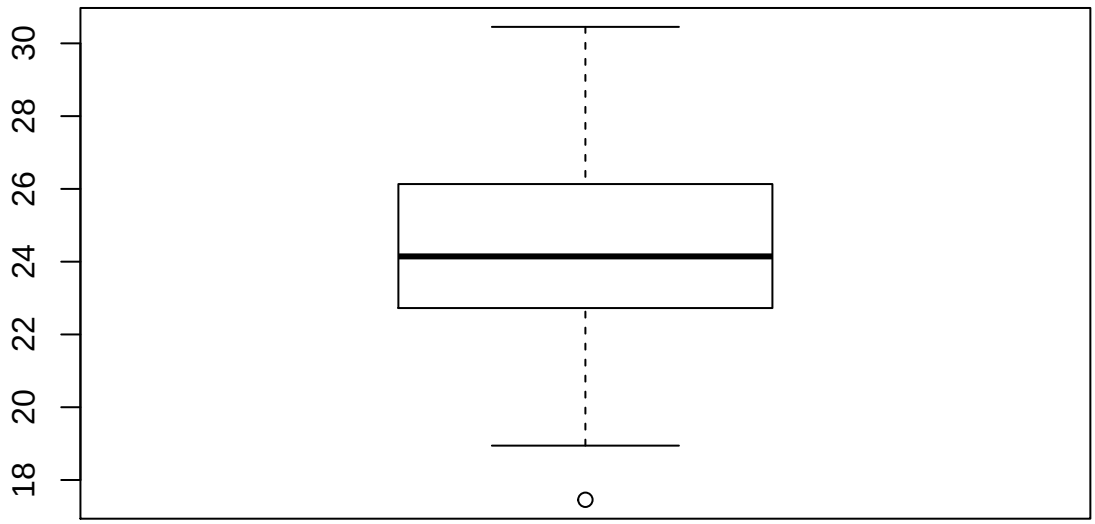
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   17.46   22.73   24.15   24.23   26.13   30.45      39
```

And we can also found that there are 39 missing values in log GDP.

Part c

In addition, we can also take a look of box plot to check if there are any outliers or whether the log GDP is skewed:

```
boxplot(gdp$Y2013)
```



From the boxplot we can conclude that there is a outlier lies outside the range (1st quartile - 1.5 IQR, 3rd quartile + 1.5 IQR), which is the smallest number. And the data is slightly skewed to the left (i.e. negatively skewed) Therefore, there is one value smaller than 1st quartile - 1.5 IQR and all the value are smaller than 3rd quartile + 1.5 IQR. We can locate the outlier using the following code:

```
gdp[which.min(gdp$Y2013),]
```

```
##     Country.Name Country.Code    Y2013
## 199       Tuvalu          TUV 17.45664
```

Hence we found that Tuvalu had the smallest GDP in 2013.

Part d

The best guess for μ and σ are $T_1 = \frac{1}{N}\sum_{i=1}^{N} Y_i$ and $T_2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$. The expected values of T_1 and T_2 are just μ and σ^2, as both of them are unbiased estimates. T_1 and T_2 and be found using the following codes.

```
mean(gdp$Y2013, na.rm = T)
```

```
## [1] 24.22638
```

```
var(gdp$Y2013, na.rm = T)
```

```
## [1] 6.093419
```