# ECON-3201 - Assignment 1

Trinh Nhat Minh Le

**Name: Trinh Nhat Minh Le**

**Student ID: 3203151**

## 1. Git and GitHub

URL to my assignment's directory on GitHub: https://github.com/minhclone4641/econ_3201.

## 2. LaTeX

### Rewrite mentioned equations in Latex

(a) $E(Y) = y_1p_1 + ... + y_kp_k = \sum_{i=1}^{k} y_ip_i$.

(b) $\sigma_Y = Var(Y) = E[(Y - \mu_y)^2] = \sum_{i=1}^{k} (y_i - \mu_y)^2 p_i$

(c) $\hat{\beta} = \frac{\sum_{i=1}^{n}(y-y_i)(x-x_i)}{\sum_{i=1}^{n}(x-x_i)^2}$

(d) $P(a \leq Y \leq b) = \int_a^b f_Y(y)dy$

(e) $\hat{g}(x) = \frac{\frac{1}{nh}\sum_{i=1}^{n} y_i k(\frac{x_i-x}{h})}{\frac{1}{nh}\sum_{i=1}^{n} k(\frac{x_i-x}{h})}$

# 3. R

## 3.1 Assignment

### (a) Set the sample size

```
n <- 1000
```

### (b) Generate 2 random variables $u_1$ and $u_2$ With n/2 = 500 observations, ~U(0,1)

```
u1 <- runif(n/2,0,1)
u2 <- runif(n/2,0,1)
```

### (c) Generate z1 and z2

```
z1 <- sqrt(-2*log(u1))*cos(2*pi*u2)
z2 <- sqrt(-2*log(u1))*sin(2*pi*u2)
```

### (d) Generate a vector z =[z1,z2]

```
z <- c(z1,z2)
```

### (e) Generate two variables $\mu$ and $\sigma$. set $\mu = 5$ and $\sigma = 2$

```
mu <- 5
sigma <- 2
```

### (f) Generate a variable $x = \mu + \sigma x z$

```
x <- mu + sigma * z
```

### (g) Calculate the mean and standard deviation of x

```
mean(x)
```

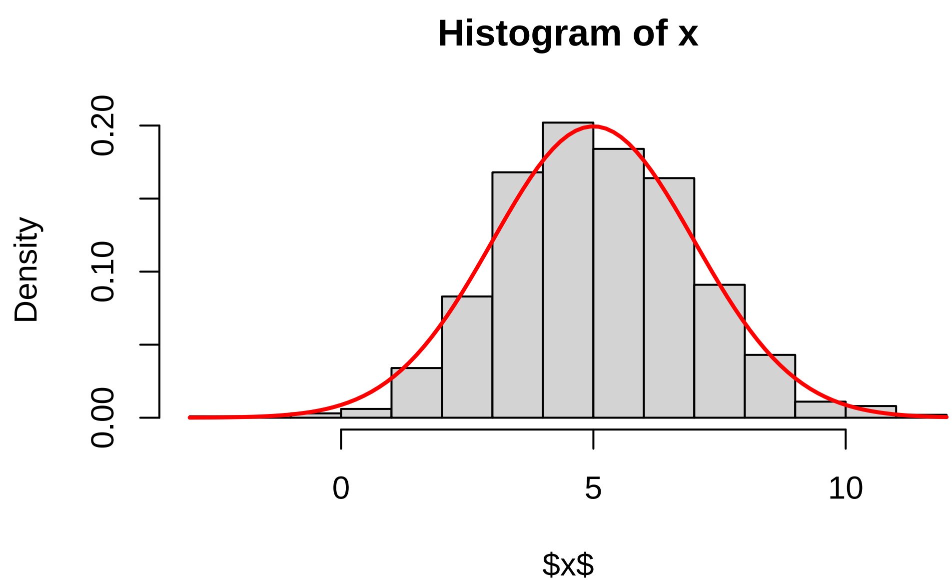
```
[1] 5.104327
```

```r
sd(x)
```

```
[1] 1.892902
```

#### (h) Plot a histogram of   with the normal distribution curve

```r
hist(x,
freq = FALSE,
ylab = "Density",
xlab = "$x$")
curve(dnorm(x, mean = mu, sd = sigma),
col = "red", lwd = 2, add = TRUE)
```




### 3.2 Data frames and Indexing

#### Load data set df from "hlthexp.csv" and adjust to match "table 1" in PDF

```r
cihi <- read.csv("hlthexp.csv")

df <- data.frame(Year = cihi$Year,
```

```
                  Hospitals = cihi$Hospitals,
                  Physicians = cihi$Physicians,
                  "Other Services" = cihi$Other.Institutions,
                  Dental = cihi$Other.Professionals..Dental.Services,
                  Vision = cihi$Other.Professionals..Vision.Care.Services,
                  "Other Professionals" = cihi$Other.Professionals..Other.Services,
                  check.names = FALSE)
```

### (a) Determine if there are any missing values for the variable Hospitals

```
any(is.na(df$Hospitals))
```

```
[1] FALSE
```

```
sum(is.na(df$Hospitals))
```

```
[1] 0
```

### (b) Add a variable called "Total Other Services" to the data frame df

```
df$'Total Other Services' <- df$Dental + df$Vision + df$"Other Professionals"
head(df)
```

```
  Year Hospitals Physicians Other Services Dental Vision Other Professionals
1 1975   5136.77    1813.15         796.62  56.40  35.86               46.72
2 1976   5977.68    2041.52         999.08  69.81  40.65               53.92
3 1977   6372.73    2252.12        1175.16  83.70  44.86               60.54
4 1978   6861.92    2528.34        1367.51 103.96  51.91               75.52
5 1979   7487.62    2804.48        1581.37 143.83  57.99               88.88
6 1980   8585.16    3235.98        1821.48 194.94  67.23              104.90
  Total Other Services
1               138.98
2               164.38
3               189.10
4               231.39
5               290.70
6               367.07
```

### (c) Skip as per instructed

**(d) Add the variable "Prescription Drugs" to the df data frame using the append method**

```
df <- data.frame(df,
                 'Prescription Drugs' = cihi$Prescribed.Drugs,
                 check.names = FALSE)
head(df)
```

```
  Year Hospitals Physicians Other Services Dental Vision Other Professionals
1 1975   5136.77    1813.15         796.62  56.40  35.86               46.72
2 1976   5977.68    2041.52         999.08  69.81  40.65               53.92
3 1977   6372.73    2252.12        1175.16  83.70  44.86               60.54
4 1978   6861.92    2528.34        1367.51 103.96  51.91               75.52
5 1979   7487.62    2804.48        1581.37 143.83  57.99               88.88
6 1980   8585.16    3235.98        1821.48 194.94  67.23              104.90
  Total Other Services Prescription Drugs
1               138.98             158.56
2               164.38             215.84
3               189.10             266.56
4               231.39             327.94
5               290.70             386.41
6               367.07             465.01
```

**(e) Using a single R command, determine the expenditure on hospitals in 1983**

```
df$Hospitals[df$Year == 1983]
```

```
[1] 13174.55
```

**(f) Using a singe R command, list the expenditures by year for 2012-2022.**

```
subset(df,
       df$Year >= 2012 & df$Year <= 2022)
```

```
   Year Hospitals Physicians Other Services  Dental Vision Other Professionals
38 2012  53299.96   29801.63       15923.80  759.13 353.62              782.67
39 2013  54954.28   31202.28       16386.15  762.36 358.08              730.08
40 2014  56123.22   32490.79       16966.03  782.00 389.71              685.88
41 2015  57352.33   33886.08       18313.73  821.42 430.46             1179.18
42 2016  58168.97   35283.98       18809.91  875.86 461.42             1355.90
```

```
43 2017  60356.12   36490.87        19665.65  918.62 484.33                  1491.51
44 2018  62896.86   37494.64        20548.31  961.17 517.89                  1614.12
45 2019  65034.33   38914.04        21446.58 1018.36 557.19                  1729.01
46 2020  67221.53   37288.46        23675.08  896.76 513.22                  1711.94
47 2021  69663.71   41479.50        25678.66  922.86 559.07                  1906.92
48 2022  73778.17   44195.30        28095.86  991.82 584.06                  2047.50
   Total Other Services Prescription Drugs
38              1895.42           12114.49
39              1850.52           12199.19
40              1857.59           12668.45
41              2431.06           13298.98
42              2693.18           13616.80
43              2894.46           13957.25
44              3093.18           14442.70
45              3304.56           14939.93
46              3121.92           15435.35
47              3388.85           16034.55
48              3623.38           17094.52
```

### 3.3. Other useful R commands.

**Load the mpg dataset from the ggplot2 package**

```r
mpg <-ggplot2::mpg
```

**(a) Subset the data to include only observations from 2008.**

```r
mpg_2008 <- subset(mpg, year == 2008)
```

**Calculate the maximum and minimum miles per gallon in city limits**

```r
#for cty's observations from 2008

max(mpg_2008$cty,na.rm = TRUE)
```

```
[1] 28
```

```r
min(mpg_2008$cty,na.rm = TRUE)
```

```
[1] 9
```

```r
#for all cty's observations

max(mpg$cty,na.rm = TRUE)
```

```
[1] 35
```

```r
min(mpg$cty,na.rm = TRUE)
```

```
[1] 9
```

#### (b) Estimate the average miles per gallon within city limits for cars produced in 2008 Using average formula

```r
sum(mpg_2008$cty)/length(mpg_2008$cty)
```

```
[1] 16.70085
```

#### (c) Estimate the average miles per gallon within city limits for cars produced in 2008 Using mean()

```r
mean(mpg_2008$cty)
```

```
[1] 16.70085
```

#### (d) Create a variable called compact

```r
mpg$compact <- ifelse(mpg$class == "compact",1,0)
```

#### (e) Estimate the average miles per gallon within city limits for compact cars.

```r
# method 1
mean(mpg$cty[mpg$compact == 1], na.rm = TRUE)
```

```
[1] 20.12766
```

```r
# method 2
mean(subset(mpg,mpg$class == "compact")$cty)
```

```
[1] 20.12766
```

**(f) Create a simple scatter plot with city mpg (cty) on the x-axis and highway mpg (hwy) on the y-axis.**

```r
plot(mpg$cty,
     mpg$hwy,

     #(i) Change the x-axis label using the option xlab = "City MPG"
     # and change the y-axis label using the option ylab = "Highway MPG"

     xlab = "City MPG",
     ylab = "Highway MPG",

     #(ii) Add the caption "City Versus Highway Fuel Efficiency (MPG)"

     main = "City Versus Highway Fuel Efficiency (MPG)")
```

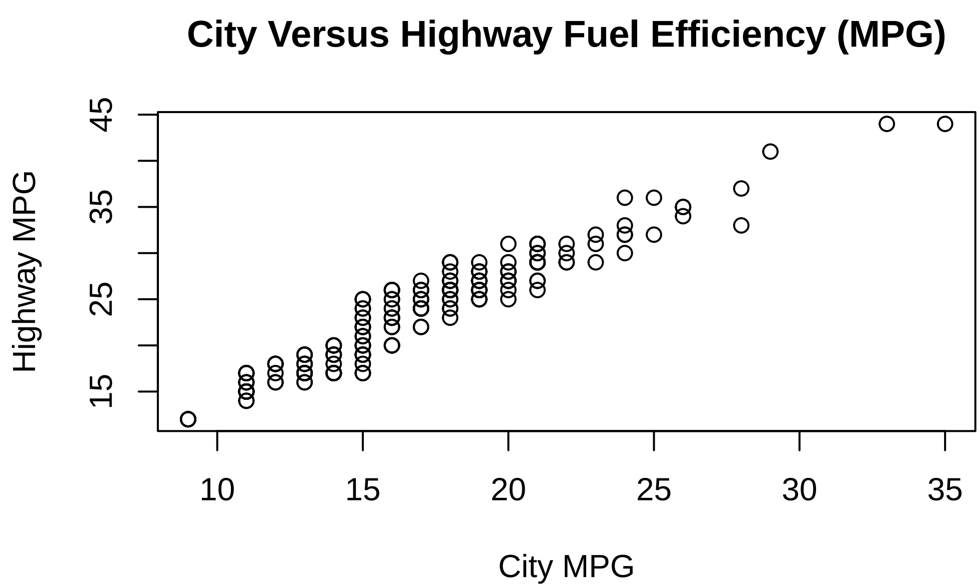



Figure 1: City Versus Highway Fuel Efficiency (MPG)

(iii) Cross reference the figure and add the text

Figure Figure 1 shows the fuel efficiency for city driving versus highway driving.