Assignment 1

2. LaTeX

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

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

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

3.1 R

(a)

```
n <- 1000
```

(b)

```
u_1 <- runif(500,0,1)
u_2 <- runif(500,0,1)
```

(c)

```
z_1 <- sqrt(-2 * log(u_1)) * cos(2 * pi * u_2)

z_2 <- sqrt(-2 * log(u_1)) * sin(2 * pi * u_2)
```

(d)

```
z <- c(z_1, z_2)
```

(e)

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

(f)

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

(g)

```
mean(x)
```

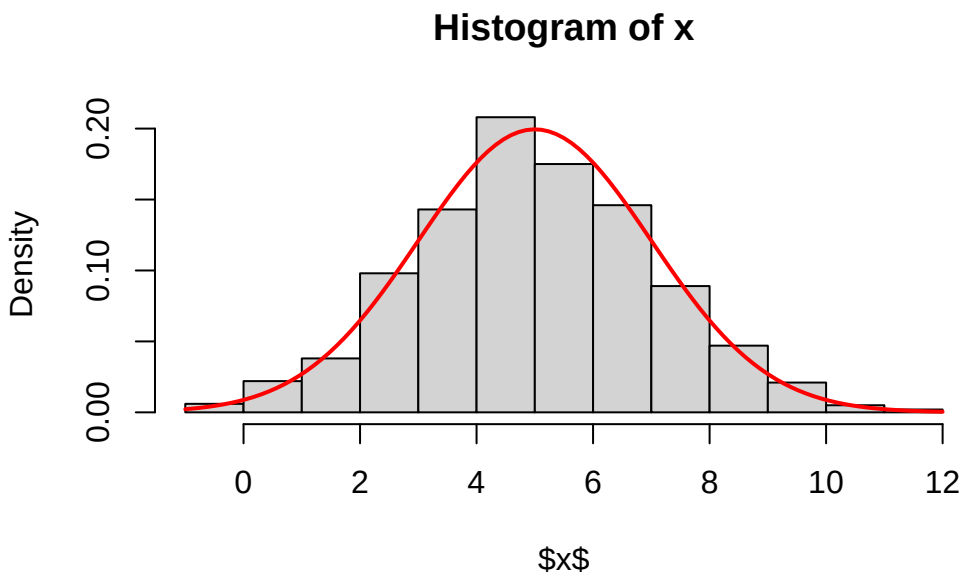
```
[1] 4.997071
```

```
sd(x)
```

```
[1] 2.054601
```

(h)

```
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 & indexing

(a) missing values from "Hospital"

```
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)
```

(b)

```
df$TotalOtherServices <- df$Dental + df$Vision + df$"Other Professionals"
```

(d)

```r
df<-data.frame(df,PrescriptionDrugs = cihi$Prescribed.Drugs)
```

(e)

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

```
[1] 13174.55
```

(f)

```r
data.frame(df$Year[df$Year>= 2012], df$Hospitals[df$Year>= 2012])
```

```
   df.Year.df.Year....2012. df.Hospitals.df.Year....2012.
1                      2012                        53299.96
2                      2013                        54954.28
3                      2014                        56123.22
4                      2015                        57352.33
5                      2016                        58168.97
6                      2017                        60356.12
7                      2018                        62896.86
8                      2019                        65034.33
9                      2020                        67221.53
10                     2021                        69663.71
11                     2022                        73778.17
```

3.3 other useful R commands

```r
#install.packages("ggplot2")
#library(ggplot2)
```

(a)

```r
mpg <- ggplot2::mpg
mpg_subset <- subset(mpg, year == 2008)
min_mpg <- min(mpg_subset$cty)
max_mpg <- max(mpg_subset$cty)
```

(b)

```
n <- length(mpg_subset$cty)
avg_mpg <- sum(mpg_subset$cty) / n
```

(c)

```
avg_mpg_mean <- mean(mpg_subset$cty)
```

(d)

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

(e)

```
avg_mpg_compact <- mean(compact)
```

(f)

```
x <- 1:10
y <- x^2
plot(mpg$cty, mpg$hwy, xlab = "City MPG", ylab = "Highway MPG", col = "blue")
```

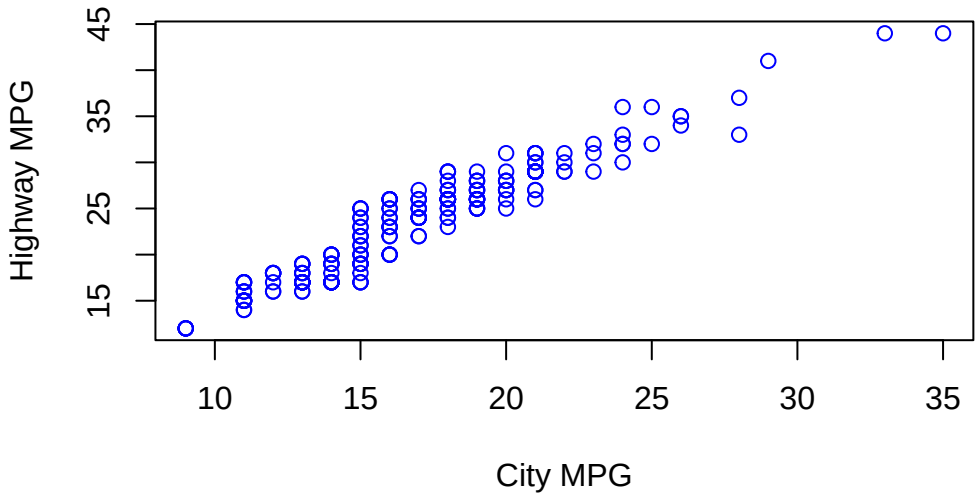


Figure 1: City Versus Highway Fuel Efficiency (MPG)

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