Sri Lanka Institute of Information Technology

Lab Submission 08

IT24102218

Chedima Imashi K.H.

Probability and Statistics - IT2120

B.Sc. (Hons) in Information Technology

1. Calculate population mean and variance of the dataset.

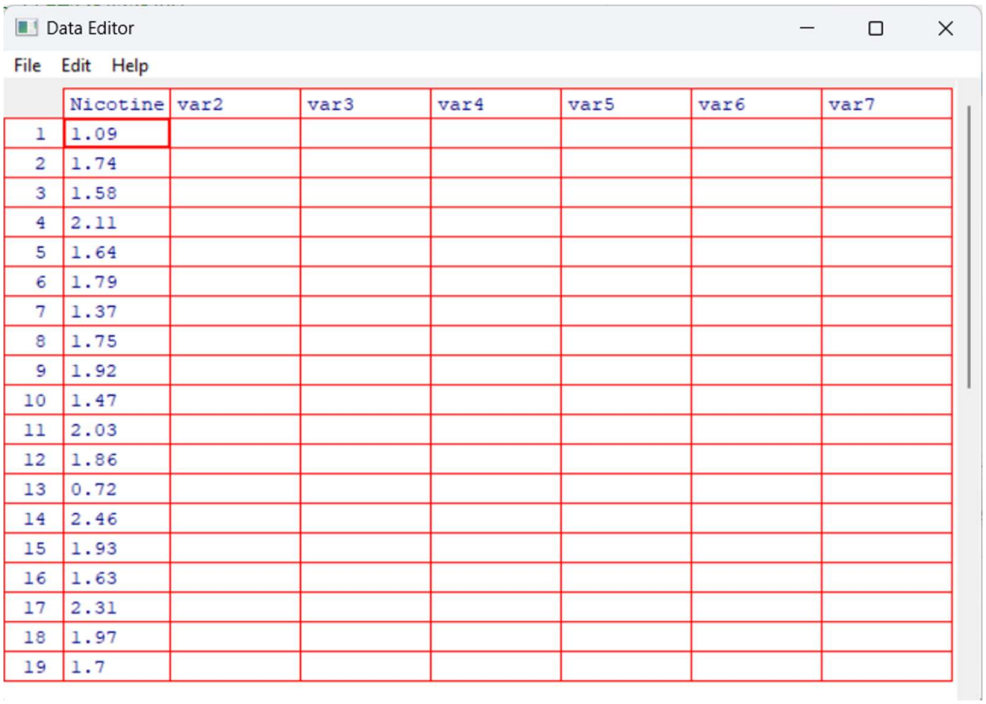
```
getwd()
setwd("C:\\Users\\User\\Desktop\\IT24102218")

data <- read.table("Data - lab 8.txt", header = TRUE)
fix(data)
attach(data)

## Q1
popmn <- mean(Nicotine)
popmn

popvar <- var(Nicotine)
popvar
```

```
> getwd()
[1] "C:/Users/User/Documents"
> setwd("C:\\Users\\User\\Desktop\\IT24102218")
> data <- read.table("Data - lab 8.txt", header = TRUE)
> fix(data)
> attach(data)
> popmn <- mean(Nicotine)
> popmn
[1] 1.77425
> popvar <- var(Nicotine)
> popvar
[1] 0.1524558
```



	Nicotine	var2	var3	var4	var5	var6	var7
1	1.09						
2	1.74						
3	1.58						
4	2.11						
5	1.64						
6	1.79						
7	1.37						
8	1.75						
9	1.92						
10	1.47						
11	2.03						
12	1.86						
13	0.72						
14	2.46						
15	1.93						
16	1.63						
17	2.31						
18	1.97						
19	1.7						

2. Get 30 random samples of size 5, with replacement and calculate sample mean and sample variance for each sample.

```
## Q2
samples <- c()
n <- c()

for(i in 1:30) {
  s <- sample(Nicotine, 5, replace = TRUE)
  samples <- cbind(samples, s)
  n <- c(n, paste('s', i))
}

colnames(samples) = n

s.means <- apply(samples, 2, mean)

s.vars <- apply(samples, 2, var)
```

```
> samples <- c()
> n <- c()
> for(i in 1:30) {
+   s <- sample(Nicotine, 5, replace = TRUE)
+   samples <- cbind(samples, s)
+   n <- c(n, paste('s', i))
+ }
> colnames(samples) = n
> s.means <- apply(samples, 2, mean)
> s.vars <- apply(samples, 2, var)
```

3. Calculate mean and variance of the Sample Means.

```
## Q3
samplemean <- mean(s.means)

samplevars <- var(s.means)
```

```
> samplemean <- mean(s.means)
> samplevars <- var(s.means)
```

4. Compare and state relationship (if any) Population Mean and the Mean of Sample Means.

```
## Q4
popmn
samplemean
```

```
> popmn
[1] 1.77425
> samplemean
[1] 1.798267
```

Population Mean : the mean of all values in the population.

Mean of Sample Means : the average of means from 30 random samples.

5. Compare and state relationship (if any) Population Variance and the Variance of Sample Means.

```
## Q5
truevar = popvar / 5
samplevars
```

```
> truevar = popvar / 5
> samplevars
[1] 0.03237593
```

Population Variance : How spread out all population values are.

Variance of Sample Means : How spread out the sample means are.

```
# Table 1: Samples 1-15
table1 <- data.frame(
  Sample = 1:15,
  Mean = round(s.means[1:15], 3),
  Variance = round(s.vars[1:15], 3)
)

print(table1, row.names = FALSE)
```

```
> # Table 1: Samples 1-15
> table1 <- data.frame(
+   Sample = 1:15,
+   Mean = round(s.means[1:15], 3),
+   Variance = round(s.vars[1:15], 3)
+ )
> print(table1, row.names = FALSE)
 Sample  Mean Variance
      1 1.976    0.120
      2 1.284    0.234
      3 1.778    0.090
      4 1.958    0.176
      5 1.664    0.067
      6 1.922    0.095
      7 1.936    0.070
      8 1.776    0.187
      9 1.634    0.103
     10 1.928    0.127
     11 1.806    0.026
     12 2.108    0.063
     13 1.736    0.145
     14 1.858    0.028
     15 2.022    0.075
```

```r
# Table 2: Samples 16-30
table2 <- data.frame(
  Sample = 16:30,
  Mean = round(s.means[16:30], 3),
  Variance = round(s.vars[16:30], 3)
)

print(table2, row.names = FALSE)
```

```
> # Table 2: Samples 16-30
> table2 <- data.frame(
+   Sample = 16:30,
+   Mean = round(s.means[16:30], 3),
+   Variance = round(s.vars[16:30], 3)
+ )
> print(table2, row.names = FALSE)
 Sample  Mean Variance
     16 1.960    0.206
     17 1.804    0.038
     18 1.538    0.306
     19 1.766    0.255
     20 1.474    0.470
     21 1.784    0.196
     22 1.898    0.115
     23 1.490    0.226
     24 1.816    0.038
     25 1.818    0.207
     26 1.998    0.052
     27 1.692    0.014
     28 1.790    0.231
     29 1.860    0.157
     30 1.874    0.022
```

```r
# Summary table
summary_table <- data.frame(
  Metric = c("Population Mean", "Population Variance",
             "Mean of Sample Means", "Variance of Sample Means"),
  Value = c(round(popmn, 3), round(popvar, 3),
            round(samplemean, 3), round(samplevars, 3))
)

print(summary_table, row.names = FALSE)
```

```
> # Summary table
> summary_table <- data.frame(
+   Metric = c("Population Mean", "Population Variance",
+              "Mean of Sample Means", "Variance of Sample Means"),
+   Value = c(round(popmn, 3), round(popvar, 3),
+             round(samplemean, 3), round(samplevars, 3))
+ )
> print(summary_table, row.names = FALSE)
                   Metric Value
          Population Mean 1.774
      Population Variance 0.152
     Mean of Sample Means 1.798
 Variance of Sample Means 0.032
```