# Faculty of Computing

## Year 2 Semester 1 (2025)

IT2120 - Probability and Statistics Lab Sheet 08

### Example Questions

```
> setwd("C:\\Users\\Victus\\OneDrive\\Desktop\\IT24102426")
>
> #Example
> #Importing the data set
> data<-read.table("Data - Lab 8.txt", header=TRUE)
> fix(data)
> attach(data)
```

```
The following object is masked from data (pos = 3):

    Nicotine
```

Data Editor

File Edit Help

| | Nicotine | var2 | var3 | var4 | var5 | var6 |
|---|---|---|---|---|---|---|
| 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 | | | | | |
| 20 | 1.9 | | | | | |
| 21 | 1.69 | | | | | |
| 22 | 1.88 | | | | | |
| 23 | 1.4 | | | | | |
| 24 | 2.37 | | | | | |
| 25 | 1.79 | | | | | |
| 26 | 0.85 | | | | | |
| 27 | 2.17 | | | | | |
| 28 | 1.68 | | | | | |
| 29 | 1.85 | | | | | |
| 30 | 2.08 | | | | | |
| 31 | 1.64 | | | | | |
| 32 | 1.75 | | | | | |
| 33 | 2.28 | | | | | |
| 34 | 1.24 | | | | | |
| 35 | 2.55 | | | | | |
| 36 | 1.51 | | | | | |
| 37 | 1.82 | | | | | |

1)

```
> #Q1
> #"mean" & "var" will compute the mean and variance for data.
> popmn<-mean(Nicotine)
> print(popmn)
[1] 1.77425
> popvar<-var(Nicotine)
> print(popvar)
[1] 0.1524558
```

2)

```
> #Q2
> #create null vectors to store sample data sets.
> samples<-c()
> n<-c()
> print(samples)
NULL
> print(n)
NULL
> #"for" loop will be used to create and assign samples of size 5 for "samples" variable created above
> #"sample" command we can draw a random sample either with replacement or without replacement.
> #By making "replace" argument as TRUE we can create samples with replacement.
> for(i in 1:30){
+   s<-sample(Nicotine,5,replace=TRUE)
+   samples<-cbind(samples,s)
+   n<-c(n,paste('s',i))
+ }
> print(s)
[1] 1.86 0.85 2.55 2.46 1.90
> print(samples)
        s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s    s
[1,] 1.58 1.79 2.55 0.85 1.85 1.75 1.79 1.64 1.85 1.68 1.75 1.92 1.47 1.64 1.75 1.69 1.88 2.09 1.85 1.64 2.46 2.17 1.85
[2,] 1.90 1.63 1.70 1.74 1.69 1.51 1.86 0.72 1.85 1.75 2.55 1.63 1.92 1.74 2.11 1.58 1.24 0.72 2.46 2.46 1.93 0.72 2.31
[3,] 2.31 2.37 1.47 2.08 2.55 1.69 1.88 1.82 1.40 1.92 1.68 1.86 1.88 2.11 1.70 2.31 1.24 2.46 1.09 1.37 1.69 2.55 1.64
[4,] 1.24 1.93 2.37 1.70 1.64 2.28 1.93 2.11 1.88 1.40 2.11 2.31 2.37 1.90 1.75 1.86 2.17 1.40 1.67 1.79 2.31 2.28 1.64
[5,] 1.37 1.93 2.03 1.63 1.97 1.63 2.11 1.75 1.92 2.08 1.51 1.74 2.55 0.72 1.90 1.68 1.40 2.28 1.85 1.70 1.85 1.63 1.09
        s    s    s    s    s    s    s
[1,] 1.75 1.69 1.37 2.09 1.74 1.79 1.86
[2,] 1.88 2.17 1.92 1.88 1.37 2.31 0.85
[3,] 1.64 2.03 1.93 1.58 1.47 1.64 2.55
[4,] 1.69 1.47 1.90 1.88 1.70 1.37 2.46
[5,] 1.75 1.82 1.64 1.51 1.75 1.51 1.90
> print(n)
 [1] "s 1"  "s 2"  "s 3"  "s 4"  "s 5"  "s 6"  "s 7"  "s 8"  "s 9"  "s 10" "s 11" "s 12" "s 13" "s 14" "s 15" "s 16"
[17] "s 17" "s 18" "s 19" "s 20" "s 21" "s 22" "s 23" "s 24" "s 25" "s 26" "s 27" "s 28" "s 29" "s 30"

> #Assign column names for each sample created. names have stored earlier under "n" variable
> colnames(samples)=n
> s.means<-apply(samples,2,mean)
> print(s.means)
  s 1   s 2   s 3   s 4   s 5   s 6   s 7   s 8   s 9  s 10  s 11  s 12  s 13  s 14  s 15  s 16  s 17  s 18  s 19  s 20
1.680 1.930 2.024 1.600 1.940 1.772 1.914 1.608 1.780 1.766 1.920 1.892 2.038 1.622 1.842 1.824 1.586 1.790 1.784 1.792
 s 21  s 22  s 23  s 24  s 25  s 26  s 27  s 28  s 29  s 30
2.048 1.870 1.706 1.742 1.836 1.752 1.788 1.606 1.724 1.924
> s.vars<-apply(samples,2,var)
> print(s.vars)
    s 1     s 2     s 3     s 4     s 5     s 6     s 7     s 8     s 9    s 10    s 11    s 12    s 13    s 14    s 15
0.18625 0.07580 0.20208 0.20585 0.13340 0.08852 0.01453 0.27667 0.04595 0.06598 0.17190 0.06707 0.18347 0.28582 0.02807
   s 16    s 17    s 18    s 19    s 20    s 21    s 22    s 23    s 24    s 25    s 26    s 27    s 28    s 29    s 30
0.08393 0.17538 0.51900 0.24008 0.16397 0.10492 0.52515 0.19343 0.00807 0.07618 0.06007 0.05717 0.03043 0.13148 0.45933
```

3)

```
> #Q3
> #"samplemean" & "samplevars" will compute the mean and variance of sample
means stored in "s.means" variable.
> samplemean<-mean(s.means)
> print(samplemean)
[1] 1.803333
> samplevars<-var(s.means)
> print(samplevars)
[1] 0.01702851
```

4)

```
> #Q4
> #compare the population mean and mean of the sample means
> popmn
[1] 1.77425
> samplemean
[1] 1.803333
```

5)

```
> #Q5
> #compare the population variance and variance of sample means
> popvar
[1] 0.1524558
> truevar=popvar/5
> samplevars
[1] 0.01702851
```

Environment History Connections Tutorial

Import Dataset 147 MiB List

R Global Environment

| | |
|---|---|
| Data | |
| data | 40 obs. of 1 variable |
| $ Nicotine: num | 1.09 1.74 1.58 2.11 1.64 1.79 1.37 1.75 1.92 1.47 ... |
| samples | num [1:5, 1:30] 1.58 1.9 2.31 1.24 1.37 1.79 1.63 2.37 1... |
| Values | |
| i | 30L |
| n | chr [1:30] "s 1" "s 2" "s 3" "s 4" "s 5" "s 6" "s 7" "s 8" ... |
| popmn | 1.77425 |
| popvar | 0.152455833333333 |
| s | num [1:5] 1.86 0.85 2.55 2.46 1.9 |
| s.means | Named num [1:30] 1.68 1.93 2.02 1.6 1.94 ... |
| s.vars | Named num [1:30] 0.1862 0.0758 0.2021 0.2059 0.1334 ... |
| samplemean | 1.80333333333333 |
| samplevars | 0.0170285057471264 |
| truevar | 0.0304911666666667 |

Exercise

```
[1] 1.570987
> setwd("C:\\Users\\Victus\\OneDrive\\Desktop\\IT24102426")
>
> #Exercise
> #Importing the data set
> data<-read.table("Exercise - LaptopsWeights.txt", header=TRUE)
> fix(data)
> attach(data)

 The following object is masked from data (pos = 3):

     Weight.kg.

> # Extract the weight column
> laptop_bag_weights <- data$Weight
```

Data Editor

File Edit Help

| | Weight.kg. | var2 | var3 | var4 | var5 | var6 | var7 | var8 |
|---|---|---|---|---|---|---|---|---|
| 1 | 2.46 | | | | | | | |
| 2 | 2.45 | | | | | | | |
| 3 | 2.47 | | | | | | | |
| 4 | 2.71 | | | | | | | |
| 5 | 2.46 | | | | | | | |
| 6 | 2.05 | | | | | | | |
| 7 | 2.6 | | | | | | | |
| 8 | 2.42 | | | | | | | |
| 9 | 2.43 | | | | | | | |
| 10 | 2.53 | | | | | | | |
| 11 | 2.57 | | | | | | | |
| 12 | 2.85 | | | | | | | |
| 13 | 2.7 | | | | | | | |
| 14 | 2.53 | | | | | | | |
| 15 | 2.28 | | | | | | | |
| 16 | 2.2 | | | | | | | |
| 17 | 2.57 | | | | | | | |
| 18 | 2.89 | | | | | | | |
| 19 | 2.51 | | | | | | | |
| 20 | 2.47 | | | | | | | |
| 21 | 2.66 | | | | | | | |
| 22 | 2.06 | | | | | | | |
| 23 | 2.41 | | | | | | | |
| 24 | 2.65 | | | | | | | |
| 25 | 2.76 | | | | | | | |
| 26 | 2.43 | | | | | | | |
| 27 | 2.61 | | | | | | | |
| 28 | 2.57 | | | | | | | |
| 29 | 2.73 | | | | | | | |
| 30 | 2.17 | | | | | | | |
| 31 | 2.67 | | | | | | | |
| 32 | 2.05 | | | | | | | |
| 33 | 1.71 | | | | | | | |
| 34 | 2.32 | | | | | | | |
| 35 | 2.23 | | | | | | | |
| 36 | 2.76 | | | | | | | |
| 37 | 2.7 | | | | | | | |

1)

```
> #Q1
> # Question 1: Population Mean and Standard Deviation
> popmn <- mean(Weight.kg.)
> popmn
[1] 2.468
> popsd <- sd(Weight.kg.)
> popsd
[1] 0.2561069
```

2)

```
> # Extract the weight column
> laptop_bag_weights <- data$Weight
>
> #Q1
> # Question 1: Population Mean and Standard Deviation
> popmn <- mean(Weight.kg.)
> popmn
[1] 2.468
> popsd <- sd(Weight.kg.)
> popsd
[1] 0.2561069
> #Q2
> #create null vectors to store sample data sets.
> samples<-c()
> n<-c()
> # Generate 25 random samples of size 6 (with replacement)
> for (i in 1:25) {
+   s <- sample(laptop_bag_weights, 6, replace = TRUE)  # one sample
+   samples <- cbind(samples, s) # add column
+   n <- c(n, paste('s', i))
+ }
>
> colnames(samples) <- n  # assign column names
>
> # Calculate sample means and variances column-wise
> s.means <- apply(samples, 2, mean)
> s.means
     s 1      s 2      s 3      s 4      s 5      s 6      s 7      s 8      s 9     s 10     s 11     s 12 
s 13     s 14     s 15     s 16     s 17     s 18     s 19     s 20 
2.466667 2.463333 2.493333 2.493333 2.566667 2.315000 2.446667 2.410000 2.673333 2.606667 2.451667 2.383333 
2.676667 2.250000 2.425000 2.531667 2.575000 2.566667 2.418333 2.495000 
    s 21     s 22     s 23     s 24     s 25 
2.555000 2.450000 2.568333 2.421667 2.573333 

> s.vars  <- apply(samples, 2, var)
> s.vars
        s 1         s 2         s 3         s 4         s 5         s 6         s 7         s 8         s 9 
s 10        s 11        s 12        s 13        s 14        s 15 
0.021306667 0.074946667 0.072706667 0.088426667 0.031906667 0.122910000 0.191266667 0.049760000 0.008666667 
0.089386667 0.057576667 0.080026667 0.008146667 0.064120000 0.039230000 
       s 16        s 17        s 18        s 19        s 20        s 21        s 22        s 23        s 24 
s 25 
0.083616667 0.085270000 0.013306667 0.102376667 0.032230000 0.061510000 0.050760000 0.027736667 0.064696667 
0.012666667 
> s.sds   <- apply(samples, 2, sd)
> s.sds
       s 1        s 2        s 3        s 4        s 5        s 6        s 7        s 8        s 9       s 10 
s 11       s 12       s 13       s 14       s 15       s 16 
0.14596803 0.27376389 0.26964174 0.29736622 0.17862437 0.35058523 0.43734045 0.22306950 0.09309493 0.29897603 
0.23995138 0.28288985 0.09025889 0.25321927 0.19806565 0.28916547 
      s 17       s 18       s 19       s 20       s 21       s 22       s 23       s 24       s 25 
0.29201027 0.11535453 0.31996354 0.17952716 0.24801210 0.22529980 0.16654329 0.25435539 0.11254629 
>
```

```
>
> #Make a table of results
> samples_df<- data.frame(
+   Sample = colnames(samples),
+   Sample_Mean = round(s.means,3),
+   Sample_SD    = round(s.sds,3)
+ )

>
> print("=== Sample means, variances,standard deviation(25 samples) ===")
[1] "=== Sample means, variances,standard deviation(25 samples) ==="
> print(samples_df)
     Sample Sample_Mean Sample_SD
s 1     s 1       2.467     0.146
s 2     s 2       2.463     0.274
s 3     s 3       2.493     0.270
s 4     s 4       2.493     0.297
s 5     s 5       2.567     0.179
s 6     s 6       2.315     0.351
s 7     s 7       2.447     0.437
s 8     s 8       2.410     0.223
s 9     s 9       2.673     0.093
s 10   s 10       2.607     0.299
s 11   s 11       2.452     0.240
s 12   s 12       2.383     0.283
s 13   s 13       2.677     0.090
s 14   s 14       2.250     0.253
s 15   s 15       2.425     0.198
s 16   s 16       2.532     0.289
s 17   s 17       2.575     0.292
s 18   s 18       2.567     0.115
s 19   s 19       2.418     0.320
s 20   s 20       2.495     0.180
s 21   s 21       2.555     0.248
s 22   s 22       2.450     0.225
s 23   s 23       2.568     0.167
s 24   s 24       2.422     0.254
s 25   s 25       2.573     0.113
```

3)

```
> #Q3
> # Calculate mean of the sample means
> samplemean <- mean(s.means)
> samplemean
[1] 2.491067
> # Calculate variance of the sample means
> samplevars <- var(s.means)
> samplevars
[1] 0.01026432
> # Standard deviation of sample means
> samplesd <- sd(s.means)
> samplesd
[1] 0.101313
>
```

```
> ## True population mean
> #compare the population mean and mean of the sample means
> #sample means ≈ population mean (The sample mean is an unbiased estimator of the population mean)
> popmn
[1] 2.468
> samplemean
[1] 2.491067
>
> # Theoretical SD of sample means
> popsd
[1] 1.570987
> theoretical_sd=popsd/sqrt(6)
> theoretical_sd
[1] 0.6413527
>
> # Population standard deviation (true σ)
> popsd <- sqrt(popvar)
> popsd
[1] 1.570987
>
> #Relationship
> print("Relationship:")
[1] "Relationship:"
> print("1. The mean of the sample means is approximately equal to the population mean (Law of Large Number
s).")
[1] "1. The mean of the sample means is approximately equal to the population mean (Law of Large Numbers)."
> print("2. The standard deviation of the sample means is smaller than the population SD and approximately eq
uals population SD divided by sqrt(sample size) (Central Limit Theorem).")
[1] "2. The standard deviation of the sample means is smaller than the population SD and approximately equals
population SD divided by sqrt(sample size) (Central Limit Theorem)."
> print(paste("Theoretical SD of sample means =", round(popsd/sqrt(6),3)))
[1] "Theoretical SD of sample means = 0.641"
```

**Environment** | **History** | **Connections** | **Tutorial**

Import Dataset · 142 MiB · List

R · Global Environment

| Data | |
|---|---|
| data | 40 obs. of 1 variable |
| samples | num [1:6, 1:25] 2.45 2.23 2.42 2.47 2.66 2.57 2.17 2.75 2.3... |
| samples_df | 25 obs. of 3 variables |
| **Values** | |
| i | 25L |
| laptop_bag_weights | num [1:40] 2.46 2.45 2.47 2.71 2.46 2.05 2.6 2.42 2.43 2.53 ... |
| n | chr [1:25] "s 1" "s 2" "s 3" "s 4" "s 5" "s 6" "s 7" "s 8" "s... |
| popmn | 2.468 |
| popsd | 1.57098695093244 |
| popvar | 2.468 |
| s | num [1:6] 2.75 2.57 2.51 2.6 2.6 2.41 |
| s.means | Named num [1:25] 2.47 2.46 2.49 2.49 2.57 ... |
| s.sds | Named num [1:25] 0.146 0.274 0.27 0.297 0.179 ... |
| s.vars | Named num [1:25] 0.0213 0.0749 0.0727 0.0884 0.0319 ... |
| samplemean | 2.49106666666667 |
| samplesd | 0.101313000518562 |
| samplevars | 0.0102643240740741 |
| theoretical_sd | 0.641352737059205 |
| truevar | 0.411333333333333 |