

CONTINUOUS ASSESSMENT – 4

ADVANCED STATISTICS

R PROGRAMS

Module 1 :

Binomial distribution :

```
#Binomial distribution
dbinom(4,size=5,prob=0.5) #for x=4
dbinom(5,size=5,prob=0.5) #for x=5
dbinom(x=2, size=5, prob = 1/2)
```

```
[1] 0.3125
```

```
[Done] exited with code=0 in 0.303 seconds
```

```
[Running] Rscript "c:\Users\Shriram kp\Desktop\Desktop files\R\R basic\tempCodeRunnerFile.r"
```

```
[1] 0.03125
```

```
[Done] exited with code=0 in 0.296 seconds
```

```
[Running] Rscript "c:\Users\Shriram kp\Desktop\Desktop files\R\R basic\tempCodeRunnerFile.r"
```

```
[1] 0.15625
```

Linear Regression : (MODULE 2)

```
#linear regression :
Advertise_money = c(1.7,1.5,2.8,5,1.3,2.2,1.3)
Monthly_sales = c(368,340,665,954,331,556,376)
relation = lm(Advertise_money~Monthly_sales) #linear model using ~
summary(relation)
result = data.frame(Advertise_money = 2.2) #for 10 system predicted 85.3 marks
final = predict(relation,result)
final
plot(Advertise_money,Monthly_sales)
```

```
Call:
lm(formula = Advertise_money ~ Monthly_sales)

Residuals:
    1      2      3      4      5      6      7 
0.25495 0.21192 -0.31008 0.26974 0.06238 -0.29901 -0.18990

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.6180149  0.2797988  -2.209 0.078224 .
Monthly_sales  0.0056062  0.0005033  11.139 0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2857 on 5 degrees of freedom
Multiple R-squared:  0.9613,    Adjusted R-squared:  0.9535 
F-statistic: 124.1 on 1 and 5 DF,  p-value: 0.0001017

Warning message:
'newdata' had 1 row but variables found have 7 rows
    1      2      3      4      5      6      7 
1.445051 1.288079 3.110080 4.730259 1.237623 2.499008 1.489900
```

Confusion matrix : (MODULE 3)

```
#Import required library
library(caret)

#Creates vectors having data points
expected_value <- factor(c(1,0,1,0,1,1,1,0,0,1))
predicted_value <- factor(c(1,0,0,1,1,1,0,0,0,1))

#Creating confusion matrix
example <- confusionMatrix(data=predicted_value, reference = expected_value)

#Display results
example
```

```
Loading required package: lattice
Loading required package: ggplot2
Warning message:
package 'caret' was built under R version 4.0.5
Confusion Matrix and Statistics
```

```
          Reference
Prediction 0 1
          0 3 2
          1 1 4
```

```
          Accuracy : 0.7
          95% CI : (0.3475, 0.9333)
No Information Rate : 0.6
P-Value [Acc > NIR] : 0.3823
```

```
          Kappa : 0.4
```

```
McNemar's Test P-Value : 1.0000
```

```
          Sensitivity : 0.7500
          Specificity : 0.6667
Pos Pred Value : 0.6000
Neg Pred Value : 0.8000
Prevalence : 0.4000
Detection Rate : 0.3000
Detection Prevalence : 0.5000
Balanced Accuracy : 0.7083
```

```
'Positive' Class : 0
```

Module 4 :

F Test

```
#Ftest  
  
A = c(8,5,7,8,3,2,7,6,5,7)  
B = c(3,7,5,6,5,4,4,5,3,6)  
# var test in R  
var.test(A, B, alternative = "two.sided")
```

F test to compare two variances

data: A and B

F = 2.4103, num df = 9, denom df = 9, p-value = 0.2061

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

0.5986736 9.7036782

sample estimates:

ratio of variances

2.410256

Module 5

K Medoids.

```
#k medoids
library(cluster)
library(factoextra)
meddata <- read.csv("kmedoidssum.csv")
fit<- pam(x=meddata,k=2)
fit$clustering
fit$medoids

summary(fit)
fviz_cluster(fit)
```

```
[Running] Rscript "c:\Users\Shriram kp\Desktop\Desktop files\R\R basic\tempCodeRunnerFile.r"
```

```
Warning message:
package 'cluster' was built under R version 4.0.5
Loading required package: ggplot2
Welcome! Want to learn more? See two factoextra-
related books at https://goo.gl/ve3WBa
Warning message:
package 'factoextra' was built under R version 4.0.5
 [1] 1 1 1 1 1 2 2 2 2 2 2
      X X.1 X.2
[1,] 4   3   6
[2,] 8   5   3
Medoids:
      ID X X.1 X.2
[1,]  5 4   3   6
[2,]  9 8   5   3
Clustering vector:
 [1] 1 1 1 1 1 2 2 2 2 2 2
Objective function:
      build      swap
2.656456 2.422273

Numerical information per cluster:
      size max_diss av_diss diameter separation
[1,]    5 5.477226 2.938518 8.660254  3.741657
[2,]    6 3.741657 1.992069 6.480741  3.741657

Isolated clusters:
L-clusters: character(0)
L*-clusters: character(0)

Silhouette plot information:
      cluster neighbor sil_width
```

```
4      1      2  0.45870867
2      1      2  0.45625942
5      1      2  0.31716734
3      1      2  0.25333528
1      1      2 -0.06111348
9      2      1  0.64914470
8      2      1  0.62160473
10     2      1  0.57442879
11     2      1  0.46762808
7      2      1  0.46000641
6      2      1  0.35949520
```

Average silhouette width per cluster:

```
[1] 0.2848714 0.5220513
```

Average silhouette width of total data set:

```
[1] 0.4142423
```

55 dissimilarities, summarized :

| Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
|-------|---------|--------|-------|---------|-------|
| 1.414 | 3.742 | 5.385 | 5.372 | 7.071 | 9.849 |

Metric : euclidean

Number of objects : 11

Available components:

```
[1] "medoids"      "id.med"      "clustering"  "objective"   "isolation"
[6] "clusinfo"    "silinfo"     "diss"        "call"        "data"
```

Error in svd(x, nu = 0, nv = k) : infinite or missing values in 'x'

Calls: fviz_cluster -> <Anonymous> -> prcomp.default -> svd

Execution halted