SADS-II problem-11 **Date: 24-04-2024**

Perspiration from a sample of 20 healthy females was analyzed. Three components, X_1 = Sweet rate, X_2= Sodium content and X_3= Potassium content, were measured and the results, which we call the sweat data are given in the below table.

Sweat Data			
Individual	X_1 (Sweat rate)	X_2 (Sodium)	X_3 (Potassium)
1	3.7	48.5	9.3
2	5.7	65.1	8.0
3	3.8	47.2	10.9
4	3.2	53.2	12.0
5	3.1	55.5	9.7
6	4.6	36.1	7.9
7	2.4	24.8	14.0
8	7.2	33.1	7.6
9	6.7	47.4	8.5
10	5.4	54.1	11.3
11	3.9	36.9	12.7
12	4.5	58.8	12.3
13	3.5	27.8	9.8
14	4.5	40.2	8.4
15	1.5	13.5	10.1
16	8.5	56.4	7.1
17	4.5	71.6	8.2
18	6.5	52.8	10.9
19	4.1	44.1	11.2
20	5.5	40.9	9.4

Find the principal components of the above data and also find the proportion of total variation explained by first and second principal components.

Solution:-

```
> setwd("D:\\dataset\\")

> data=read.csv("principalcom.csv")
> data
     X1   X2   X3
1   3.7 48.5  9.3
2   5.7 65.1  8.0
3   3.8 47.2 10.9
4   3.2 53.2 12.0
5   3.1 55.5  9.7
6   4.6 36.1  7.9
7   2.4 24.8 14.0
8   7.2 33.1  7.6
9   6.7 47.1  8.5
10 5.4 54.1 11.3
11 3.9 36.9 12.7
12 4.5 58.8 12.3
13 3.5 27.8  9.8
```

```
14 4.5 40.2  8.4
15 1.5 13.5 10.1
16 8.5 56.4  7.1
17 4.5 71.6  8.2
18 6.5 52.8 10.9
19 4.1 44.1 11.2
20 5.5 40.9  9.4
> head(data)
    X1   X2   X3
1 3.7 48.5  9.3
2 5.7 65.1  8.0
3 3.8 47.2 10.9
4 3.2 53.2 12.0
5 3.1 55.5  9.7
6 4.6 36.1  7.9
> setwd("D:\\dataset\\")
> data=read.csv("principalcom.csv")
> head(data)
    X1   X2   X3
1 3.7 48.5  9.3
2 5.7 65.1  8.0
3 3.8 47.2 10.9
4 3.2 53.2 12.0
5 3.1 55.5  9.7
6 4.6 36.1  7.9
> ##step2: Standardize the data
> scale_data=scale(data)
> head(scale_data)
                  X1         X2         X3
[1,] -0.55396106  0.2204127 -0.3491471
[2,]  0.62467949  1.3950038 -1.0316904
[3,] -0.49502903  0.1284267  0.4909061
[4,] -0.84862119  0.5529777  1.0684427
[5,] -0.90755322  0.7157222 -0.1391338
[6,] -0.02357281 -0.6569927 -1.0841937
> ##Step 3: Perform Principal Component Analysis (PCA)
> # Perform PCA
> pca_result <- prcomp(scale_data, scale. = TRUE)
> pca_result
Standard deviations (1, .., p=3):
[1] 1.3440666 0.8955140 0.6257313

Rotation (n x k) = (3 x 3):
             PC1        PC2        PC3
X1  0.6534351 -0.1023336 -0.7500336
X2  0.4870226  0.8153574  0.3130516
X3 -0.5795097  0.5698422 -0.5826219
> # Summary of PCA
> summary(pca_result)
Importance of components:
                                  PC1    PC2
Standard deviation     1.3441 0.8955
Proportion of Variance 0.6022 0.2673
Cumulative Proportion  0.6022 0.8695
                                  PC3
Standard deviation     0.6257
Proportion of Variance 0.1305
Cumulative Proportion  1.0000
> ##Step 4: Interpret Principal Components Extract and Interpret Loadings
> # Extract loadings
> loadings <- pca_result$rotation
>
> # Print loadings
> print(loadings)
             PC1        PC2        PC3
X1  0.6534351 -0.1023336 -0.7500336
```

```
X2  0.4870226  0.8153574  0.3130516
X3 -0.5795097  0.5698422 -0.5826219
> ##step5:Proportion of Variance
> # Proportion of variance explained
> prop_variance <- pca_result$sdev^2 / sum(pca_result$sdev^2)
>
> # Print proportion of variance
> print(prop_variance)
[1] 0.6021717 0.2673151 0.1305132
> ##step6: percentage
> per =((0.6021717+0.2673151)/sum(prop_variance))*100
> per
[1] 86.94868
```