Data Analysis Project 1
by Yiyi HUANG (20391572)

I choose the “Animal Sleeping Data”.
I performed PCA step by step and found four principal components whose accumulative contribution is larger than 0.85. The loadings show that the first of them is (strongly) positively related to variables 1, 2, 3, while (strongly) negatively related to the left. The second of them is (strongly) positively related to variables 4, 5, while (strongly) negatively related to 8, 10. The third is (strongly) negatively related to variable 1. The fourth is (strongly) positively related to variable 6.
The following are *Mathematica* codes.

In[1106]:=
```
data = Import["/mnt/sda4/Documents/data_analysis/sleep1.csv"];

mean[i_] := Mean[DeleteCases[data[[All, i]], _String]]
(* replace "NA" with mean of other data in the same column *)
dataClean =
  Transpose[Table[data[[All, i]] /. "NA" → mean[i], {i, 1, Dimensions[data][[2]]}]];
(* standardize the data *)
dataStandard = Chop[Standardize[dataClean]];
correlation = Correlation[dataStandard];
eigensystem = Eigensystem[correlation];
eigenvalue = eigensystem[[1]];
eigenvector = eigensystem[[2]];
pcaAll = dataClean.Transpose[eigenvector];
(* accumulative contribution *)
accumulativeContribution = Accumulate[eigenvalue] / Plus @@ eigenvalue
(* let accumulative contribution > 0.85, determine which components to select *)
index = Position[accumulativeContribution,
   SelectFirst[accumulativeContribution, # > .85 &]] // Flatten
(* here are the first 4 principal components *)
(* output is suppressed *)
pcaSelected = pcaAll[[All, Range[index[[1]]]]];
eigenvectorSelected = eigenvector[[Range[index[[1]]]]];
sqrtEigenvalues = Sqrt /@ eigenvalue[[Range[index[[1]]]]];
(* here are the loadings corresponding to the first 4 principal components *)
loading = Transpose[Thread[Times[eigenvectorSelected, sqrtEigenvalues]]]
```

Out[1273]=
```
{0.475991, 0.698516, 0.825389, 0.889555,
 0.937057, 0.964541, 0.983326, 0.99244, 0.997482, 1.}
```

Out[1274]=
```
{4}
```

Out[1278]=
```
{{0.70165, 0.630557, 0.846362, -0.547678, -0.619603, -0.53715,
  -0.806093, -0.557653, -0.832581, -0.72617}, {0.145444, 0.3655, 0.132938,
  0.657993, 0.705775, 0.448753, 0.411523, -0.63075, -0.201344, -0.559146},
 {-0.517922, -0.347674, -0.363877, -0.383915, -0.25312, 0.393335, -0.0501756,
  -0.455451, -0.227798, -0.345229}, {0.346339, -0.0875683, 0.277133,
  -0.260306, -0.118557, 0.526062, 0.129394, 0.0070539, 0.239899, 0.0658572}}
```