# Applications of PCA - Worksheet

**Part One**

1. What is the point of rotating principal components? What do you gain by doing so?

2. After rotation, does the first principal component/factor still explain as much variance as it did initially?

3. Since the principal components are originally decided by determining the subspace with maximum variance, wouldn't a rotation of these components explain less variance? Can you explain this?

4. *(True/False.)* The default factor analysis procedure (proc factor) in SAS provides essentially the same exact output as proc princomp (PCA).

5. *(True/False.)* The output dataset from the princomp procedure contains the eigenvectors of the correlation or covariance matrix.

6. *(True/False.)* The output dataset from the princomp procedure contains the coordinates of the observations in their new basis, which is formed by the eigenvectors of the correlation or covariance matrix.

7. *(True/False.)* Covariance and Correlation PCA are exactly the same thing.

8. The following output is produced in SAS after running the default Factor Analysis procedure on the Iris dataset.

| Factor Pattern | |
|---|---|
| | Factor1 |
| Sepal_Width | -0.46014 |
| Sepal_Length | 0.89017 |
| Petal_Width | 0.96498 |
| Petal_Length | 0.99156 |

| Variance Explained by Each Factor |
|---|
| Factor1 |
| 2.9184978 |

| Final Communality Estimates: Total = 2.918498 | | | |
|---|---|---|---|
| Sepal_Width | Sepal_Length | Petal_Width | Petal_Length |
| 0.21173131 | 0.79240043 | 0.93118439 | 0.98318168 |

a. The procedure determines one factor should be used based on the "mineigen" criteria. What is the mineigen criteria?

b. How do you interpret the communality value of 0.93 for the variable *petal width*?

c. What is the total amount of variance for this example? How do you know? If I use just this one factor to approximate my data, what is the proportion of variance that I will capture?

d. If an observation had a negative coordinate (or score) on factor1, which of the following two situations would be possible:

   i That observation had above average *sepal length, petal width, and petal length* and below average *sepal width*.

   ii. That observation had below average *sepal length, petal width, and petal length* and above average *sepal width*.

**List of Key Words/Phrases.**

loadings
eigenvector
principal components
scores
coordinates
communality
rotations
varimax rotation
proportion of variance explained
variable clustering
eigenvalues
variance explained by factors