Applications of PCA - Worksheet

Part One

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

   Rotating principal components makes our factors (new variables) more interpretable by enforcing sparsity in the factor matrix (i.e. making each factor have only a few variables with relatively high loadings)

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

   No, after rotation the first factor will not explain the same amount of variance. some of that variance will be shifted to the other components.

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?

   No, once the dimension of the subspace is decided on, the variance of the data in that lower dimensional space stays constant regardless of how I rotate the axes. While each component may not explain the same amount of variance as it did before rotation, the total variance of all the components together will remain the same

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

   TRUE

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

   FALSE

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.

   TRUE

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

   FALSE

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?

only use factors with eigenvalues greater than 1

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

using this factor model (only one factor in this case) you will account for 93 percent of the variance in the variable petal width. This is an R^2 value for predicting petal width with factor1.

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?

total variance is 4. default factor procedure uses correlation PCA. Thus, the proportion of variance would be 2.918/4 which is close to 75%

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