

Data Cleaning

Section 3: Missing Data, Outlier Detection, and Principal Component Analysis (PCA) 

LESSON 7: PRINCIPAL COMPONENT ANALYSIS (PCA)

Lesson 7: Eigenvalues

Using eigenvalues is an alternative to the scree plot. Eigenvalues are a highly advanced topic in linear algebra. They loosely associate to the importance of specific data points. An eigenvalue less than 1 is often considered trivial, as the closer the eigenvalue is to 0, the less important it is.

An eigenvalue is always a positive number greater than 0 and assigned to a component to represent how good it is at explaining variance in a data set. An eigenvalue greater than 1 means the given principal component, in a sense, is better than average at explaining variance in the data set. Therefore, the first principal component will always have an eigenvalue greater than 1. If an eigenvalue is less than 1, this principal component typically is not strong at explaining variance. This is one criterion to use when selecting the number of principal components to extract from the analysis. The larger the eigenvalue, the better the given principal component is at explaining variance.



Data Cleaning

Section 3: Missing Data, Outlier Detection, and Principal Component Analysis (PCA) ▼

LESSON 7: PRINCIPAL COMPONENT ANALYSIS (PCA)

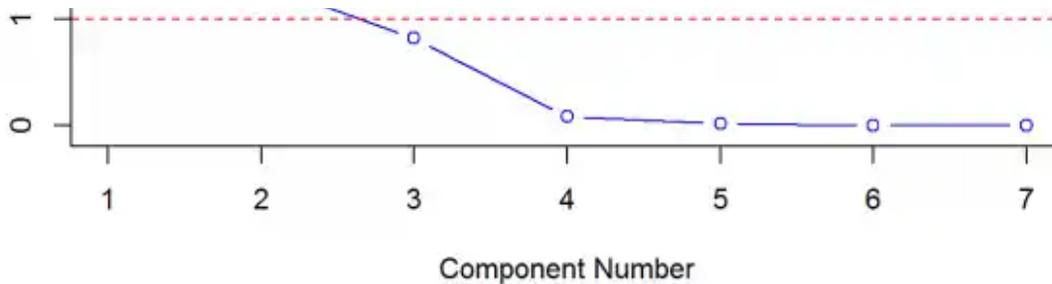


Figure: A graph showing eigenvalues with two components having values greater than 1.

In the scree plot in the figure, you should consider any components that have values greater than 1 for the eigenvalue. There are two components that meet this criteria. However, in the example below, four components meet this criteria.

Data Cleaning

Section 3: Missing Data, Outlier Detection, and Principal Component Analysis (PCA) ▼

LESSON 7: PRINCIPAL COMPONENT ANALYSIS (PCA)

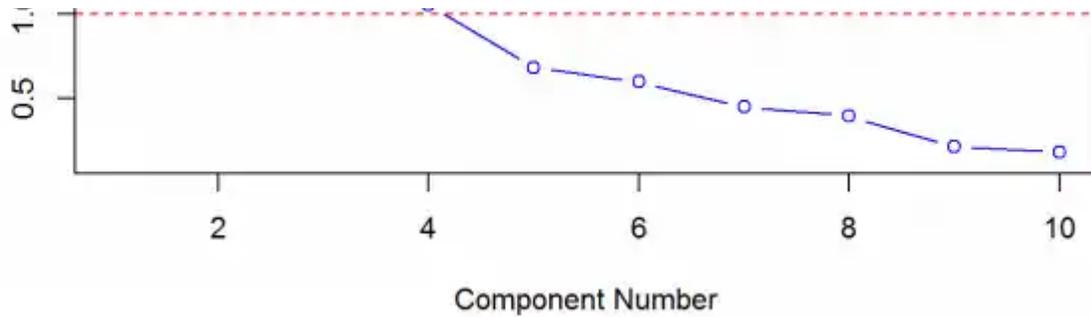


Figure: A graph showing eigenvalues with four components having values greater than 1.

