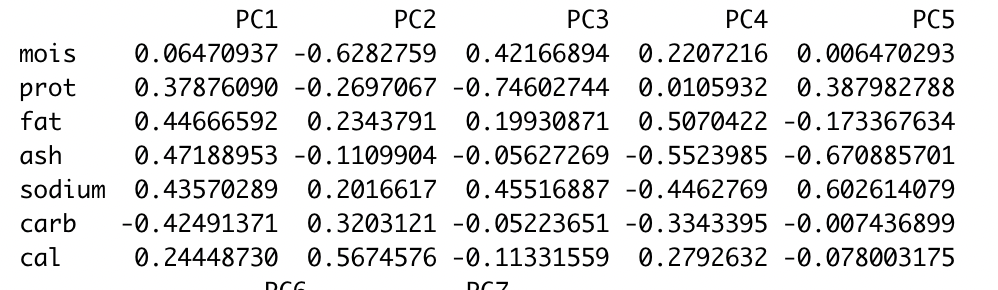
**Do you have so many variables in your dataset(s) that you just don’t know where to start?**

Then Principal Component Analysis (PCA), a dimension reduction technique, could be useful for you! Analysis can help identify patterns in a data set and distill the variables down to their most prominent features so that the data is simplified without losing much of the data’s variability.

|  |  |
| --- | --- |
| **Data Requirements for PCA**   * Data for the predictor variables must be numeric * No outcome variable * Some correlation (linear relationship) between predictor variables * Three or more independent predictor variables * Standardized or scaled (although PCA can help with this) | By transforming the independent data, PCA highlights the maximum variances in the dataset. It does so by standardizing the data and creating a matrix that summarizes how variables all relate to each other. Then PCA computes the eigenvectors and assesses the magnitude and direction of them. The eigenvectors with the largest eigenvalues will become the principal components (PCs). These PCs convey the strongest correlations between all the variables. The first principal component has the most variance, second principal component has the second most variance, and so forth. |

Sample PCA Output:



Columns indicate principal components.

Rows indicate variables and their correlations within the components.

The direction and magnitude of the principal components are the most important for recognizing trends in the data set. The further a value is from zero, the more the variable contributes to the component. In the output above, fat, ash, sodium and carb variables have the largest magnitudes in PC1. Therefore, we can conclude that fat, ash and sodium will all jointly increase, while carb will decrease when fat, ash, and sodium increase.

Ted Guevel, Teresa Vail, Kayla Strunk

**Recommended Reading**

1. Interpretation of Principal Components <https://online.stat.psu.edu/stat505/lesson/11/11.4>
2. Overview <https://towardsdatascience.com/a-one-stop-shop-for-principal-component-analysis-5582fb7e0a9c>
3. PCA Practical Guide in R & Python <https://www.analyticsvidhya.com/blog/2016/03/pca-practical-guide-principal-component-analysis-python/>
4. Step By Step Video [StatQuest: Principal Component Analysis (PCA), Step-by-Step](https://www.youtube.com/watch?v=FgakZw6K1QQ)
5. Step by Step Explanation Principal Component Analysis <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>
6. Overview and Explanation <https://www.sartorius.com/en/knowledge/science-snippets/what-is-principal-component-analysis-pca-and-how-it-is-used-507186>
7. Principal Component Analysis Explained for Dummies <https://programmathically.com/principal-components-analysis-explained-for-dummies/>
8. PCA Analysis in R <https://www.datacamp.com/tutorial/pca-analysis-r>
9. Principal Components Analysis in R <https://www.statology.org/principal-components-analysis-in-r/>
10. Principal Component Analysis PCA in R <https://www.r-bloggers.com/2021/05/principal-component-analysis-pca-in-r/>