Principal Components Analysis

DATA: Planets.sav
We want to classify planets - which planets are similar to one another?
We measure 6 characteristics (rings, moons, dsun, radius, mass, density).
Why are some planets more alike than others? How many dimensions underlie the data

A1. Convert measurements to logarithms (make data linear)
A2. Principal Components Analysis
ROTATION: Varimax
DESCRIPTIVES: Correlation Coefficients
SCORES: Save as variables; Display factor scores
EXTRACTION: Principal Components; Scree Plot
A3. 2 dimensions account for 96% of the total variance in our 6 measurements
A4. Rotated component matrix (mass, radius, rings, moons vs. distance and density)
SCATTERPLOT: Factor1 vs. Factor2
LABEL CASES BY: Planet
A5. Mercury, Venus, Earth, Mars are all similar
A6. Uranus, Neptune, Saturn, Jupiter are all similar
A7. Inner rocky planets, outer gas giants, and pluto
A8. We already knew this -- a simple example

Cluster Analysis

DATA: Planets.sav
We want to group planets by their similarity?
To do this, we'll calculate distances between planets (in terms of their variable scores)

A1. Standardize variables (put all on the same scale)
A2. Cluster Analysis
CLASSIFY: Hierarchical Cluster
STATISTICS: Proximity Matrix
PLOTS: Dendogram (Horizontal)
METHOD: Between Groups Linkage
CLUSTER: Cases
DISPLAY: Statistics & Plots
A3. Look at dendogram (we can have 1-8 clusters)
A4. Look at icicle. If we want to have 3 groups…

Cluster Analysis – A realistic application

DATA: Nations.sav
We want to group nations by their similarity?
Based on 12 measures of living-conditions

A1. We need to standardize the variables (GNP has a huge range, life expectancy is small)
A2. Standardize, range of 1, other methods
A2. Cluster Analysis
CLASSIFY: Hierarchical Cluster
STATISTICS: Proximity Matrix, Range of 2-6 clusters
PLOTS: Dendogram (Horizontal); Range of 2-6 clusters
METHOD: Centroid Clustering
A3. Look at dendogram (we can have 1-8 clusters)