

Name \_\_\_\_\_

*Psychology 516*  
*Applied Multivariate Analysis*  
Homework 9  
Due November 6, 2018

The file, Set\_7\_A.csv, contains data for 183 participants in a Pew Center Political Survey conducted in April 2017. Participants were asked their opinions about government spending using the following question stem:

*“If you were making up the budget for the federal government this year, would you increase spending, decrease spending or keep spending the same for \_\_\_\_\_”?*

The spending areas considered were:

- a) Scientific research
- b) Military defense
- c) Government assistance for the unemployed
- d) Medicare
- e) Environmental protection
- f) Economic assistance to needy people around the world
- g) Education

Responses were coded 1 = Decrease spending, 2 = Keep spending the same, and 3 = Increase spending.

1. First, use Ward’s procedure with Euclidean distance to determine how many clusters of respondents best describe this sample. Use the following procedures to arrive at the best choice:
  - (a) Run Ward’s procedure on a file that has been re-ordered using values for the military spending variable. Save the cluster memberships for a two-cluster solution. Repeat this procedure after re-ordering the file using values for the environmental protection variable. Repeat this procedure after re-ordering the file using values for the education variable. When you are done, you will have three new cluster membership variables, each from a Ward’s method applied to a different ordered version of the file. Create a cross-classification table for each pair of cluster membership variables. Are cases classified consistently regardless of how the data were ordered?
  - (b) Repeat this process, but for a three-cluster solution.
  - (c) Repeat this process, but for a four-cluster solution.
  - (d) Repeat this process, but for a five-cluster solution.
  - (e) Of these solutions, which one provides the highest number of clusters while being completely immune to the ordering of the variables?

The cluster memberships from each re-ordering cannot be directly compared because they have the respondents in different orders. Make sure you put the cluster memberships in a consistent order (use the ID variable) before creating the cross-classification tables.

2. For the chosen number of clusters from Question 1, create a bar graph of the cluster means for each of the spending variables (include 95% confidence intervals). Based on these graphs, provide a description of each cluster's spending attitude profile.
3. Using K-means clustering, create a plot of the within-cluster sums of squares for cluster solutions up to 10. Does the number of clusters from this plot suggest the same number of clusters identified by Ward's procedure?
4. For the chosen number of clusters from Question 1, conduct a K-means clustering of the data. Provide a cross-classification table for the Ward's and K-means procedures. Do the two procedures identify the same clusters?
5. The file, Set\_7\_B.csv, contains additional information about these respondents. Use this file to better define the nature of the clusters identified in Question 4. The variables included in Set\_7\_B are age, home area (rural, urban, suburban), level of trust in the government, gun control attitude, sex, highest level of education, race, annual income, political party affiliation, and political ideology. Using these variables, what are the defining features of each cluster? Focus on those variables that do the best job of distinguishing the groups.