STAT 4400   Final Exam

Take-home exam policy:

- You will **not** be allowed extra time. You will **not** be allowed any discussion with the classmates. Otherwise, you will receive zero for the exam.
- You will type a report detailing your analysis of the data and your conclusions. You may include graphs that illustrate and/or support your findings. Do NOT include computer code within the main body of your report. You may include such code in an appendix if you wish.
- The take home test should be turn in before 3:00pm on Wednesday, May 3.

  Thanks.

1. (30 pts) A research team has gathered data on **77 countries**, measuring 13 variables on each, which represent various characteristics of the nations: Land (in square km), Population (in millions of people), Energy (energy usage), Rural (percent of people living in rural areas), Military (spending on the military as a percentage of budget), Health (health care spending as a percentage of GDP), HIV (percent of adults who are HIV positive), Internet (percent of people with Internet access), Birth Rate, Elderly Pop (percent of population that is elderly), Life Expectancy (in years), GDP (gross domestic product), Electricity (a measure of consumption of electricity).

   The questions that we would like answered include:

   - (10 pts) Are there particular countries that are highly unusual in terms of the measured characteristics? If so, identify them. Are there notable associations/relationships between some of the variables? If so, describe them. Is there a way to graphically represent the raw data for the 77 countries and draw conclusions about the data set from such a graph?
   - (10 pts) Can we find a few indices that describe the variation in the data set using a lesser dimension than the original set of variables? If so, what are those indices? Is there a convenient interpretation of any of the indices?
   - (5 pts) Can we graphically display the data in a low number of dimensions using such indices? What conclusions about the countries (individual countries or groups of countries) can you draw from such a graph?
   - (5 pts) Are there any countries that are similar or different from each other in any aspects that are surprising to you? What useful information, e.g., for a public relations campaign, could be gleaned from this data set as related to this?

2. (15 pts) Do both a hierarchical clustering and a partitioning clustering for data **50 countries**. For each clustering, you may pick your favorite specific approach. Give the partitions of countries into clusters, give some plot(s) to visualize the cluster structure.

3. (15 pts) The following data set gives the road distances between 12 cities in a state and cities in neighboring states.
   a. (12 pts) Locate the cities in $k = 1, 2$ and 3 dimensions using multidimensional scaling.
   b. (3 pts) Compare the two-dimensional multidimensional scaling configuration with the locations of the cities on a map from an atlas or the distance matrix given.

Cities	Appleton	Beloit	Atkinson	Madison	Marshfield	Milwaukee	Monroe	Superior	Wausau	Dubuque	St.Paul	Chicago
Appleton	0											
Beloit	130	0										
Atkinson	98	33	0									
Madison	102	50	36	0								
Marshfield	103	185	164	138	0							
Milwauke	100	73	54	77	184	0						
Monroe	149	33	58	47	170	107	0					
Superior	315	377	359	330	219	394	362	0				
Wausau	91	186	166	139	45	181	186	223	0			
Dubuque	196	94	119	95	186	168	61	351	215	0		
St.Paul	257	304	287	258	161	322	289	162	175	274	0	
Chicago	186	97	113	146	276	93	130	467	275	184	395	0

4. (20 pts) The national track records for women in 54 countries can be examined for the relationships among the running events.

   a. (8 pts) Perform a principal components factor analysis on the data.
   b. (8 pts) Apply maximum likelihood factor analysis. Rotate the factor solution selected using an orthogonal procedure, and interpret the results.
   c. (4 pts) Compare principal components factor analysis to maximum likelihood factor analysis.