PH245
Introduction to Multivariate Statistics
Homework Set 4

Due date: November 20, Monday

Problems:

1. (8-18) (8-19) (8-20) The data "Data-HW4-Track-women.dat" recorded national track records for women, while the data "Data-HW4-Track-men.dat" recorded national track records for men.

   (a) Obtain the sample correlation matrix for the women track records data, and determine its eigenvalues and eigenvectors.

   (b) Determine the first two principal components for the standardized predictors. Find out the cumulative percentage of the total sample variance explained by the two components.

   (c) Interpret the two principal components.

   (d) Rank the nations based on their score on the first principal component. Does this ranking correspond with your intuitive notion of athletic excellence for the various countries?

   (e) Convert the national track records for women to speeds measured in meters per second. Notice that the records for 800m, 1500m, 3000m, and the marathon are given in minutes. The marathon is 26.2 miles, or 42,195 meters, long. Perform a principal components analysis using the covariance matrix of the speed data. Compare the results with the results in (b). Do your interpretations of the components differ? If the nations are ranked on the basis of their score on the first principal component, does the subsequent ranking differ from that in (d)? Which analysis do you prefer? Why?

   (f) Now repeat the principal component analysis outlined in (a) to (d) for the men. Are the results consistent with those obtained from the women's data?

2. (9-20) (9-21) The data "Data-HW4-Pollution.dat" recorded 42 measurements on 7 air-pollution variables recorded at 12:00 noon in the Los Angeles area on different days.

   (a) Using all 7 air-pollution variables to generate the sample covariance matrix.

(b) Obtain the principal component solution to a factor model with $m = 1$ and $m = 2$. Find the corresponding commonalities.

(c) Find the proportion of variation accounted for by the one-factor model, and the two-factor model, respectively.

(d) Perform a varimax rotation of the $m = 2$ solution, and interpret the factors after the rotation. Find the proportion of variation accounted for by the two-factor model after the rotation.

Policy: You must do the homework on your own. Please ask the Instructor or the GSI if you have any question.