STAT 730: Homework 5

1. Perform a principal components analysis on the leisure time data with the 10 variables remaining; use the *raw, unscaled* outcomes.

   ```
   d=read.csv("http://www.stat.sc.edu/~hansont/stat730/leisure.csv",
    header=T,row.names=1)
   d2=d[,3:22]
   a=c(1,2,5,6,8,9,12,15,19,20)
   colnames(d2[,a])
   f=prcomp(d2[,a])
   ```

   (a) How much variability is explained by the first two principal components?

   (b) Formally test that you can set all three of the loadings from $\hat{\boldsymbol{\gamma}}_{(1)}$ that are less than 0.15 in magnitude to zero. Use the Wald test outlined on the board in class; my original notes had the wrong $\mathbf{V}_i/n$. Formally test that you can set both of the loadings from $\hat{\boldsymbol{\gamma}}_{(2)}$ that are less than 0.15 in magnitude to zero.

   (c) Try to interpret the contrasts suggested in part (b). Note that PC1 and PC2 explain roughly the same amount of variability.

   (d) Obtain a correlation plot for the 10 variables using the first two principal components *scaling each variable to have unit variance*; this is a default plot from using `PCA` in the `FactoMineR` package. Are any of the 10 variables *mostly* explained by the first two principal components?

2. Obtain both the classical and Kruskal's MDS solutions for the beverage data (Chapter 13) based on the *standardized variables*. Which approach has lower stress, and therefore fits better according to this criterion? Will this always be the case? Do the two approaches provide similar or quite different two-dimensional maps? Note that `isoMDS` uses the classical solution as a starting point, and so the first stress reported during the iterations is the stress from the classical solution.

3. Perform canonical correlation analysis on the leisure time data. Compare the more active/social activities *PaddleRacq, TeamSp, Weights, Shop, Cook* with the more passive/solitary *Read, HikingNature, Instrument, FreeMind, TV*; this is suggested from $\hat{\boldsymbol{\gamma}}_{(2)}$ in problem 1. Try to *carefully* interpret the vectors $\mathbf{a}_1$ and $\mathbf{b}_1$. Test $H_0 : \boldsymbol{\Sigma}_{12} = \mathbf{0}$ (MKB pp. 288–289).

4. Derive the full conditional distribution for $P(z_i = j|\text{else})$ and $[\mathbf{w}|\text{else}]$ on slide 9 in the notes for Chapter 13. Note that you already essentially found the full conditionals for $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ in homework 3; here, they are updated only from the data that actually comes from component j, i.e. updated using $\mathcal{D}_j = \{\mathbf{x}_i : z_i = j\}$.