Assignment # 6, Stat 8053, Fall 2013

Reading

Class notes on MANOVA; Härdle and Simar Chapters 9-10

Problems

Due date is Monday, December 9. For these two problems, you are to provide the most complete, interesting and appropriate analysis you can based on the material on cluster analysis. For these two problems, you are to provide the most complete, interesting and appropriate analysis you can based on the material, stressing factor analysis and cluster analysis.

1. Use the swiss dataset:

> library(MASS)

> data(swiss)

Information about the variables can be obtained from ?swiss. Apply both factor analysis and clustering in a sensible fashion, and summarize results.

2. The data file below was collected from the US Statistical Abstract for 2012 and for 2006. The data includes crime rates per 100,000 population in 7 categories of crimes, for both 2009 and for 2005. Four of the categories, murder, rape, robbery and assault, are personal crimes and the other three are property crimes. In a few cases there are missing values, mostly because data are collected in those cities in a different format, so I suggest you simply delete those cities. In the following questions, use clustering and any methods we have learned that you might find helpful to answer interesting questions based on these data. For clustering, you might want to look at larger cities

```
> load(url("http://z.umn.edu/crimedatarda"))
> crimedata <- na.omit(crimedata)</pre>
```

- (a) What does factor analysis and cluster analysis suggest about crime rates in 2005? Are the 2005 results consistent with the 2009 data? What has changed?
- (b) Compute the change in crime rates within cities over the two time periods (you might look at differences, ratios, or percent change, 100(2009 rate 2005 rate)/2005 rate). Apply factor analysis and/or clustering again to learn about how crime has changed over these four years.
- (c) Use cluster analysis to cluster crime rates rather than cluster cities, and summarize results.