**Group 20-Chunhui Lin, Menghan Lin, Xinyu Wu**

**Topic: Canonical Correlation Analysis**

**Data: National Health and Nutrition Examination Survey 2003-2004**

**Language: R(Xinyu Wu), Stata (Chunhui Lin), SAS(Menghan Lin)**

**Brief introduction:**

Canonical correlation analysis is used to identity and measure the associations among two sets of variables. In our project, we focus on the relationship between personal exposure to 10 volatile organic compounds and biochemical liver tests. Serum albumin, total bilirubin, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, alkaline phosphatase and y-glutamyl transferase served as the outcome variables. Personal exposures to benzene, chloroform, ethylbenzene, tetrachloroethene, toluene, trichloroethene, o-xylene, m-,p-xylene, 1,4-dichlorobenzene, and methyl tert-butyl ether (MTBE) are assessed as another set of interest.

To sum up, set one is volatile organic compounds (VOCs) includes:

benzene, chloroform, ethylbenzene, tetrachloroethene, toluene, trichloroethene, o-xylene, m-,p-xylene, 1,4-dichlorobenzene, and methyl tert-butyl ether (MTBE)

Another set of variables is status of biochemical liver contains:

Serum albumin, total bilirubin, alanine aminotransferase, aspartate aminotransferase, lactate dehydrogenase, alkaline phosphatase and y-glutamyl transferase

**Citation:**

Liu, J., Drane, W., Liu, X., & Wu, T. (2009). Examination of the relationships between environmental exposures to volatile organic compounds and biochemical liver tests: Application of canonical correlation analysis. *Environmental Research,109*(2), 193-199. doi:10.1016/j.envres.2008.11.002

https://www.cnblogs.com/pinard/p/6288716.html

https://onlinecourses.science.psu.edu/stat505/node/63/

http://users.stat.umn.edu/~helwig/notes/cancor-Notes.pdf