Remy Lagrois

Experimental Statistics

Project 3

Cluster Analysis of Osteoporosis in Women Over 55

**Introduction**

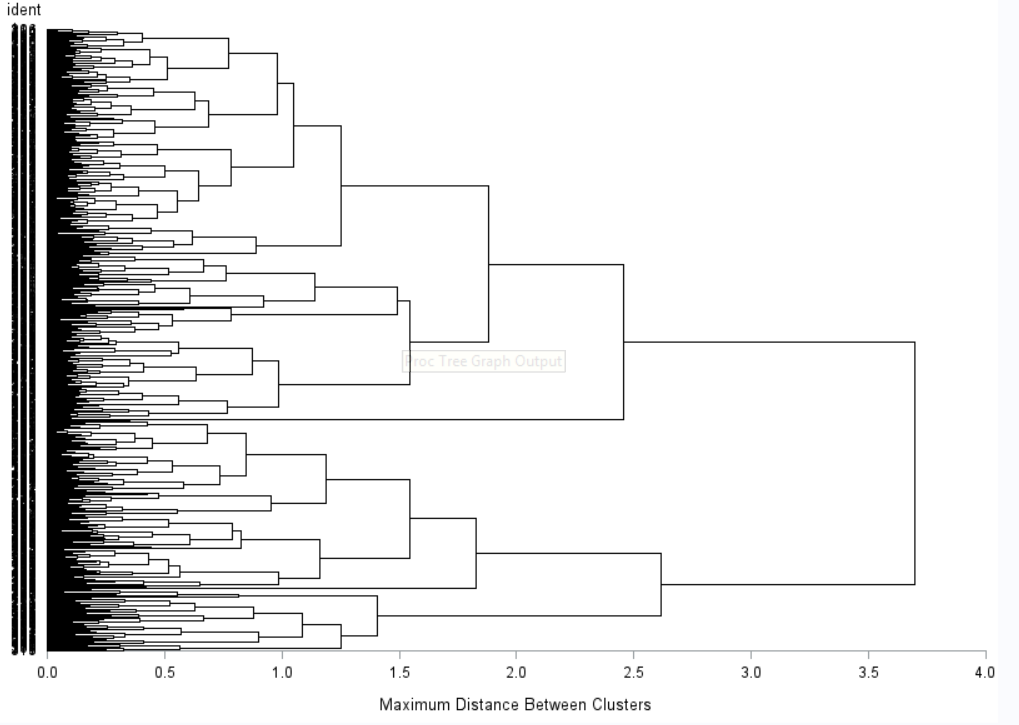
As any person gets older their chances of fracturing a bone increase as their bone density decreases (i.e. loss of calcium). This problem is especially apparent in women. As a woman moves out of reproductive age she will go through the process of menopause. While the primary effect of menopause is the halting of menstruation, it also includes an increased risk of osteoporosis.

The Global Longitudinal study of Osteoporosis in Women (GLOW) is, as the name suggests, a study looking at osteoporosis, risk factors, and outcomes in women over the age of 55. The primary goals are to identify the primary risk factors and reduce fractures. This is important both on the personal and national level. Breaking a bone is not a positive experience at any age but as a person gets older the consequences of a fracture (and the ability to recover) become worse and worse. This also can have economic impacts as the person or their family are prevented from working and causing an increase in insurance premiums more widely. Reducing bone fractures in the elderly will mean fewer people having to go through the experience and reduced costs (literally) for everyone else.

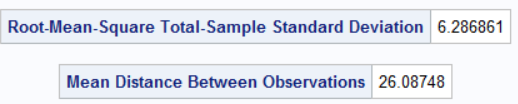
In this paper we perform a cluster analysis on five-hundred GLOW observations. We use complete linkage to try to find previously hidden groupings of women and their (potential) fractures. We then ran an analysis of variance on the clusters to find differences between them in terms of the means of our variables.

**The Data**

As previously mentioned we used five-hundred observations to run the analysis. The GLOW researchers collected the following information about each woman in the study: ID, study site, physician ID, history of fractures (hist), age of enrollment (age), weight at enrollment (wt), height at enrollment (ht), BMI (bmi), menopause before or after 45 (meno\_45), mother’s hip fractured or not (mom\_frac), use of arms needed when standing or not (arm\_ast), smoker or not, self-reported risk (risk), risk score (frac\_score), and any fractures in the past year (frac). For our cluster analysis we did not include any of the IDs as they don’t have anything to do with risk of fracture and we also did not include smoking information since there were so few smokers in the study and its removal improved our clusters.

**Analysis** 

Above is the dendrogram resulting from running a cluster analysis using complete linkage and including the variables as described above. The value of ‘nclusters’ was set to five. Specifying a higher number resulted in several clusters containing only one or two observations while selecting fewer resulted in indistinct clusters while retaining clusters of one or two observations. As there are 500 observations it is difficult to pick out where each individual is. However, we can see that the data is split into two main groupings at a distance between 3.5 and 4. That first split is somewhat uneven with slightly more than 60% falling into the top group.



The standard deviation of the total-sample root-mean-square came out to 6.29 while the mean distance is 26.1. Using the other linking algorithms resulted in standard deviations of up to around 8 and mean distances of up to around 29.