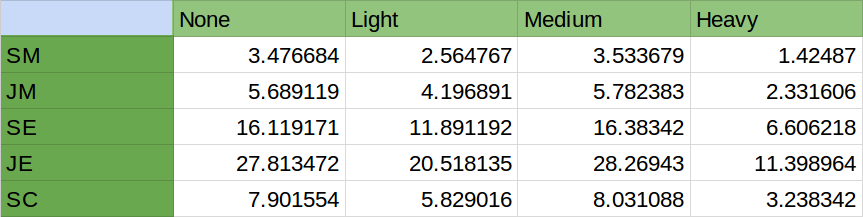
# [MS-E2112 - Multivariate statistical analysis](https://mycourses.aalto.fi/course/view.php?id=7621) – Home exercise 5

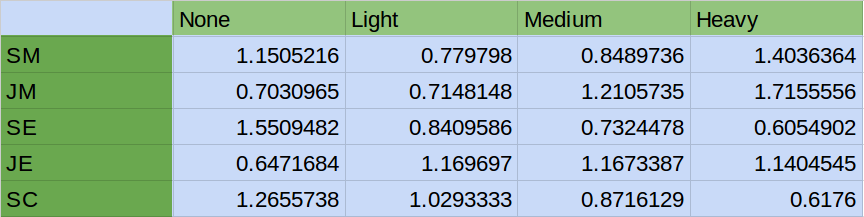
**Problem 1**

**Data**: A 2-dimensional frequency table SMOKING.txt data was provided, where the details of smoking of employees in a company was tabulated. The employees were categorized into Senior Managers (SM), Junior managers (JM), Senior Employee (SE), Junior Employee (JE), Secretaries (SC), which were further categorized into None, Light, Medium and Heavy smokers.

a) The theoretical frequencies for the given data under independence is given in the table below. Each element is the product of its corresponding row and columns’ **marginal relative frequency**. But the margin remain same as the observed frequencies.



b) The attraction repulsion matrix for the data is given below. It tells about the relationship between two modalities. If two modalities have a value greater than 1, it means they are attracted, less than 1, they are repulsed (unattracted) and if 1, it means they not related.



c) From the attraction repulsion matrix given above, the following analysis can be made.

1. Junior Managers are the most frequent smokers (here, in heavy category) of all and Secretaries are least frequent smokers (also in heavy category!).
2. The smoking group of Junior Employees is most frequent of all the groups this is also highlighted by the fact they are repulsed in ‘None’ category with 0.64, denoting most of them are smokers.
3. The Senior Employees are least frequent smokers which is also substantiated by more frequency in ‘None’ category.

**Appendix**

The code for the problem solved above.

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

setwd("~/Documents/OneDrive/Aalto/Sem2/MSA/Ex 5")

data <- read.table("SMOKING.txt",header = T, row.names = 1)

data <- data[,-5]

data <- data[-6,]

D <- as.matrix(data)

# standardizing the row and columns so that they add to 1

prop.table(D,1) #row profile

prop.table(D,2) #col profile

v1 <- margin.table(D,1) # Gives you the sum of all the rows

v2 <- margin.table(D,2) # Gives you the sum of all columns

n1 = length(v1)

n2 = length(v2)

V1 = matrix(v1,ncol = 1)

V2 = matrix(v2,ncol=n2)

E = V1 %\*% V2 / sum(D)

AR.matrix <- D/E # D = original data (number of observations)

# E = expected number of observations under independence

round(AR.matrix,2) # Round off to two digits

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%