Factor Analysis – Reduce number of variables and lower multicollinearity

Factors are linear combination of the variables

F1 = A11X1+A12X2+A13X3+A14X4…….+A1NXN

F2= A21X1+A22X2+…….+A2NXN

….

FN = AN1X1+AN2X2+…..+ANNXN

Aij – Coefficient of jth Variable in the ith Factor (factor Loadings)

Factors are rotated in an N dimensional Plane till they become mutually orthogonal, i.e. Perpendicular and hence independent with each other

Consider those factors which account for 90-95% of the total variation in the data. For these selected factors we consider the variables with high loadings. This variable set is used for further analysis.

100 Variables – 10 Factors which account for 90-95% variation

– 15 Variables with high loading (will have lower multicollinearity than the original data) are chosen from these 10 factors

– Principal Component Analysis

PROC FACTOR:

NFACT: NO OF FACTORS GENERALLY KEPT SAME AS THE NUMBER OF VARIABLES

METHOD: PRIN (SHORT FORM FOR PRINCIPAL COMPONENT ANALYSIS)

ROTATE : VARIMAX (ROTATION METHOD ENSURING ORTHOGANILITY OF FACTORS)

OUT : OUTPUT FILE WHICH STORE THE FACTOR SCORES