PRINCIPAL COMPONENTS

Here is principal components in base R.

> result <- prcomp(dbvars,center = TRUE, scale =TRUE)

> summary(result)

Importance of components:

PC1 PC2 PC3 PC4 PC5

Standard deviation 1.2015 1.0453 1.0172 0.9385 0.7405

Proportion of Variance 0.2887 0.2185 0.2070 0.1762 0.1097

Cumulative Proportion 0.2887 0.5072 0.7142 0.8903 1.0000

> result$rotation

PC1 PC2 PC3 PC4 PC5

age 0.52382885 0.3419017 0.3417354 -0.50742307 -0.4841955

fem -0.21739388 -0.5248529 0.7039190 0.23527334 -0.3555472

sales -0.52583690 0.1674794 0.3842556 -0.59335723 0.4424037

seniority 0.63378319 -0.3098845 0.2856979 0.01925758 0.6483026

singapore -0.01255619 0.6953811 0.3980443 0.57855391 0.1520644

Note that, while social scientists expect to see eigenvalues, R gives standard deviations of the principal components.

Moreover, instead of printing factor loadings (the correlations between the principal components and each original variable), it provides “rotations”, which are raw eigenvector values and do not equal the correlation of the variables with the components.

See next page for Stata output.

Stata command and output:

. factor age fem sales seniority singapore, pcf

Factor analysis/correlation Number of obs = 258

Method: principal-component factors Retained factors = 3

Rotation: (unrotated) Number of params = 10

--------------------------------------------------------------------------

Factor | Eigenvalue Difference Proportion Cumulative

-------------+------------------------------------------------------------

Factor1 | 1.44355 0.35098 0.2887 0.2887

Factor2 | 1.09257 0.05779 0.2185 0.5072

Factor3 | 1.03479 0.15399 0.2070 0.7142

Factor4 | 0.88080 0.33250 0.1762 0.8903

Factor5 | 0.54829 . 0.1097 1.0000

--------------------------------------------------------------------------

LR test: independent vs. saturated: chi2(10) = 60.82 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

-----------------------------------------------------------

Variable | Factor1 Factor2 Factor3 | Uniqueness

-------------+------------------------------+--------------

age | 0.6294 0.3574 0.3476 | 0.3553

fem | -0.2612 -0.5486 0.7161 | 0.1181

sales | -0.6318 0.1751 0.3909 | 0.4174

seniority | 0.7615 -0.3239 0.2906 | 0.2308

singapore | -0.0151 0.7269 0.4049 | 0.3075

-----------------------------------------------------------

Note that it prints real eigenvalues, and real factor loadings.

See next page for my bpca function

Here is my bpca() function:

> bpca(dbvars)

===============================================

PCA RESULTS (Stata/SPSS Style Output)

===============================================

EIGENVALUES AND VARIANCE EXPLAINED

-----------------------------------

Component Eigenvalue Variance\_Pct Cumulative\_Pct

PC1 1.4436 28.87 28.87

PC2 1.0926 21.85 50.72

PC3 1.0348 20.70 71.42

PC4 0.8808 17.62 89.03

PC5 0.5483 10.97 100.00

Minimum eigenvalue threshold (mineigen): 1

Components retained: 3

COMPONENT LOADINGS (Variable-Component Correlations)

----------------------------------------------------

PC1 PC2 PC3

age 0.629 0.357 0.348

fem -0.261 -0.549 0.716

sales -0.632 0.175 0.391

seniority 0.761 -0.324 0.291

singapore -0.015 0.727 0.405

COMMUNALITIES (Variance Explained per Variable)

-----------------------------------------------

Variable Communality

age 0.645

fem 0.882

sales 0.583

seniority 0.769

singapore 0.692