

Homework 7:

Principal Components Analysis (PCA)

You will reuse your data file (all 32 variables) from homework 6, part b. You will now practice the use of PCA to illuminate the structure of this same data set. Your write-up (excluding the summary page) must not exceed 8 pages.

1. Conduct a principal component analysis (PCA). In the first run, let the program extract as many components as it wants. For SPSS, the commands look like this:

```
FACTOR VAR = DISTANT TO EASYGOIN  
  /plot = eigen  
  /format = sort  
  /extr = pc  
  /rota = varimax.
```

Note: I used the command /format = sort in the above syntax. It orders the items in such a way as to group those together that load highest on a particular PC. This greatly improves the interpretability of the component (loading) matrices.]

2. (a) Inspect the scree plot and comment on the strength (size of eigenvalue) of the various components.
(b) Why did the program extract the number of components it did?
(c) Does the scree plot support the number of components that the program extracted?
3. Look at the unrotated factor loading (“component”) matrix. Can you find approximate interpretations for each component? (See again the Big Five table, posted under Homework 5) for background information.) You don’t need to adopt a particular theoretical stance here; just write a sentence or two about a reasonable interpretation (if possible) of each component.
4. Now look at the rotated loading matrix (by default SPSS uses VARIMAX as the rotation algorithm, which aims at simple structure).
 - (a) Describe to what degree there is a “simple structure.”
 - (b) Try to interpret the components you see. Find a meaningful label and describe the variables that hang together under that label.
 - (c) What happened to the variances of each principal component after rotation?
5. Researchers often use the results of a PCA as a guideline for forming subscales or unidimensional “tests.” Your task is to create a reliable scale based on the *third* rotated component.
 - (a) Compute or ask SPSS to save this component “C3” so that it is available as a variable in your data file.

- (b) What is the variance of C3?
 - (c) Now re-inspect the loading matrix and *select* those original variables that load highly on this third component. Use a reasonable selection cut-off (there is no single correct one) and compute the aggregate of those variables' standard scores, showing mean and variance of the resulting variate. Let's call it "C3unit" because it uses unit weights ($w = 1$) for each of its variables.
 - (d) Compare C3 (original third component computed from precise weights) and C3unit, using statistical and distributional parameters and a bivariate graph. What differences, if any, do you notice?
 - (e) Using SPSS's RELIABILITY procedure, compute Cronbach's α for the C3unit variable. Comment on the quality and parsimony of the scale. Perform item deletion, if useful, to improve its α .
6. Write a one-page summary, reporting on the PCA results (unrotated and rotated) and, briefly, on the comparison of the variates C3 and C3unit.