

**University of North Texas
Department of Geography and the Environment**

GEOG 5190

LAB 4: FACTOR ANALYSIS

On the course website you have been provided with a file that contains demographic data for Denton County's 137 census tracts.

The 15 fields recorded in this file for each census tract in Denton County are, in order:

Tract:	the tract code from the US Census Bureau
County:	the county code from the US Census Bureau
State:	the state code from the US Census Bureau
Young:	the percent of the census tract population that falls in the 0-4 years age group
Old:	the percent of the census tract population that falls in the 65 and over age group
Household:	the average household size in the census tract
Income:	the median income of households in the census tract
Afr-Amer:	the percent of the population that is African-American
Hispanic:	the percent of the population that is Hispanic
Married:	the percent of families in the census tract that have legally-married spouses
HU_Owner:	the percent of dwellings in the census tract that are occupied by their owners
HU_Vacant:	the percent of dwellings in the census tract that are vacant
Employed:	the percent of the age 18 and over population that is employed
Exp_Ind:	index reflecting household expenditures relative to average (US average = 100)
Crime_Ind:	index reflecting crime levels relative to national average (US average = 100)

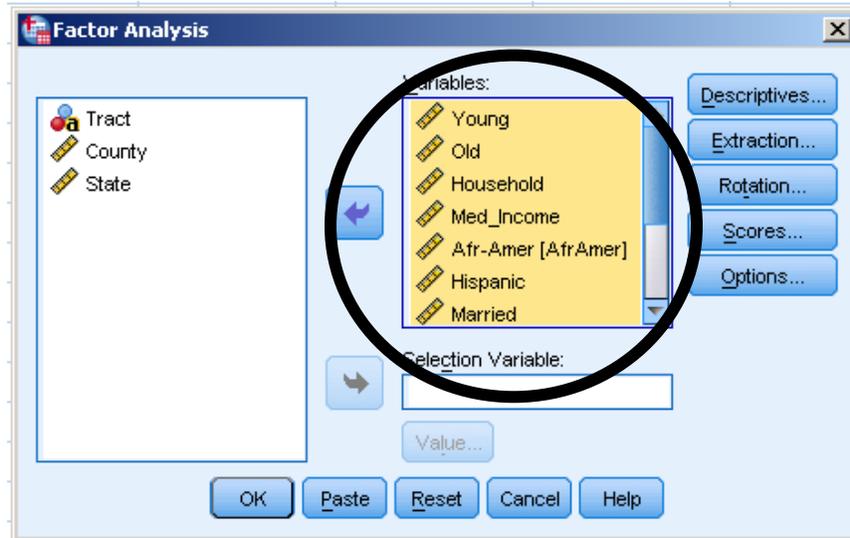
Out of these 15 fields:

- The first 3 are identifier fields that define the geography we are dealing with (tract/county/state)
- The next 12 are data fields that give us some insight into what the population of each census tract is like.

It is these 12 data fields that form the focus for this lab analysis.

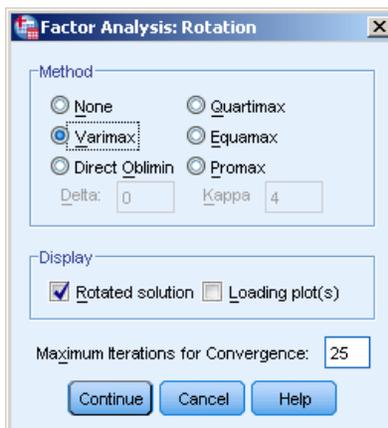
Your task here is to use factor analysis to simplify this situation. The central question: even though we have 12 data fields (actual tract information) that make up a fairly complex database, can we pull out of the database a smaller number of overall factors that characterize much of what is going on in the database?

In SPSS, go to **Analyze > Dimension Reduction** to reach the **Factor...** (Factor Analysis) module. Once in the factor analysis dialogue window, you'll see some variable lists and buttons that are similar to what you've already used in previous labs (see graphic below). Enter all of the variables, other than "Tract", "County" and "State", into the factor analysis.



Use the **Rotation...** button in the right portion of the factor analysis window (see above) to set the kind of rotation used.

Use two rotations for this lab: first, use the Varimax rotation, then for your next analysis, use another rotation of your choosing. As part of the documentation for your report, **name the other rotation you chose and describe it briefly** (in other words, discuss briefly the goal of this rotation type). Use the "Help" button in the Factor Analysis dialogue box shown below for further information about the rotations available through SPSS (the help function contains a brief description of each). **Note:** you will need to answer the questions on the following page for both of the rotation analyses you do.



Once you have selected your rotation method, click "Continue" in the rotation dialogue above, then click "OK" in the main Factor Analysis dialogue to have SPSS complete your analysis.

For each of the two rotations you use, answer the four questions below:

Question 1. How many factors does SPSS find *initially* (based on “initial eigenvalues”) and *after rotation* (under “Rotation Sums of Squared Loadings”).

Question 2. Find the “Total Variance Explained” (Rotation Sums of Squared Loadings) table. How much of the total variation in the database does factor 1 account for? How much of the total variation does the last factor account for on its own? What do all of the factors account for (cumulative percent)?

Question 3. Find the “Rotated Component Matrix” table. This is the “factor matrix” we discussed in our PC/FA lecture. Recall that the numbers in this table are our “factor loadings” that help us interpret each factor. Using the information in this table, name each of the components (factors) based on the variables that *load high* for each in a *positive direction* (here we’ll use those +0.6 and higher) and in a *negative direction* (we’ll use -0.6 and lower).

- For example, if one of the factors were to have a loading value of +0.9 for our “Young” variable, we could say that a strong part of that particular factor would be the presence of young children.
- For another example, if one of the factors were to have a loading value of -0.8 for our “HU_Vacant” variable, we could say that a strong part of that factor would be the absence of vacant housing (or, to put it another way, a high presence of occupied housing).

Question 4. What insights do you gain from the factor analysis? In other words, describe what the factors tell us about the database: how has doing this factor analysis enhanced your understanding of this dataset?

Answer this last question once (after you’ve run both rotations):

Question 5. Are there differences between the two rotations you have used? If so, what are they?

For all of the above:

Make sure you give careful thought to the database, the situation, and the methods you’re using so your answers truly reflect what is important to know here.

**Please remember to include a copy of your SPSS output
in your written report.**

**Feel free to also integrate selected tables from the SPSS output in
your answers above if you think that could be helpful.**