Exercise

Exercise A

1. Select the three indicators with the largest raw headcounts. Tabulate them by pairs. For each pair, compute the Cramer V and the P Measure. What do you conclude when comparing the two measures.

Exercise B

1. Examining the weight structure in the 'Wealth Index'

In this exercise we will explore how the PCA is used as an ad-hoc data reduction technique to generate aggregated indices. We will discuss the advantages and disadvantages of this family of techniques as an aggregation method and as a procedure to set weights.

Data: '2. OPHI-HDCA SS11 - Working Groups on Principal Components and Factor Analyses, data ss_haiti06, JMR.dta.dta'

Variables to consider in the index:
-   Dichotomous variables: radio television refrigerator bicycle motorbike car telephone electricity roof floor,
-   Ordinal scales: water cookingfuel toilet wall (the categories of these variable are also available as dummys variables in case you prefer to use these)
-   Continue: crowding

Sample weight: weight

1.1   Compute a wealth index using PCA and using the dichotomous variables (this is the method followed by Filmer and Pritchett 2001). If you feel confident, run the analysis with the matrix of polychoric correlations as in Kolenikov (2009). See commands at the bottom of this document.

Use the command:

.factor varlist [if] [in] [weight] [, method options] ➜ use the method pcf

Save the score from the first component with the command:

.predict [type] newvar [if] [in] [, single_options]

1.2   Observe the factor loading and discuss the rational of the final weight and whether they conflict with more some normative criteria. Do you agree or disagree?

2. Explore weight structure based on PCA and using your indicators

Please repeat run a PCA with your own indicators as in exercise 1. Interpret the weights as above for your indicators.

Basic commands in STATA to implement Exploratory Factor Analysis

```
.factor varlist [if] [in] [weight] [, method options ]
     Method:           pf       principal factor; the default
                       pcf   principal-component factor
                       ipf       iterated principal factor
                       ml   maximum-likelihood factor


     Useful options: factors(#)   maximum number of factors to be
                     retained (usually, the analysis is run for the
                     first time without setting a number of
                     factors)
                     blanks(#) display loadings as blanks when
                     |loadings| < # (this is useful but it is also
                     relevant that the analyst observes all
                     loadings in order to identify variable that
                     might be contributing with more than one
                     factor simultaneously).


. screeplot [eigvals] [, options ]


. findit fapara (in case the ADO file was not installed yet)
. fapara, reps(10)


.  rotate [, options]
     Useful options:  varimax  executes the method "varimax" during
                     the rotation; this is the most common
                     orthogonal method
                     promax executes the method "promax" during the
                     rotation; this is the most common oblique
                     method
                     blanks(#)  display loadings as blank when
                     |loadings| < #; default is blanks(0)


 .predict [type] newvar [if] [in] [, single_options]



If the variables are dummies, the analysis can be implemented with
the tetrachoric correlations matrix, by using the following
commands:


. tetrachoric varlist [if] [in] [weight] [, options]
. matrix r = r(R)
.factormat r, n(#) {where # is the number of cases considered
```

Similarly, polychoric correlations matrix might be used if the variables are ordinal with few categories, to assume cardinality (it is also useful for mixed variables – dummies, ordinal and continuous variables). If the number of the categories of one of the variables is greater than 10, polychoric treats it is continuous, so the correlation of two variables that have 10 categories each would be simply the usual Pearson moment correlation found through the command correlate

```
.findit polychoric
. polychoric varlist [if] [in] [weight] [, options]
. matrix r = r(R)
.factormat r, n(#) {where # is the number of cases considered
```

It is worth noticing that the estimation of polychoric correlations is computationally demanding and that, under certain circumstances (sufficient categories for some type of analysis), the results are equivalent to those of the standard method, as mentioned previously.