# Lab Handout for Homework #2: FOR STATA VERSION 9.0

In this handout, we briefly cover the commands needed to perform factor analysis in stata version 9 and some commands associated with problem set2. You would encounter two situations when performing factor analysis: (1) with variables in the dataset; (2) with correlation matrix (as part (5) in problem set2.) Details are described below:

## (1) Principal components analysis: performed to decide how many factors should be retained

### (1.1) for variables

| Description | Command | default | Preceded by | Followed by | Example: |
|---|---|---|---|---|---|
| **Principal components analysis** | **pca** | | | llist of variables | pca a b c |
| **Specifies what type of matrix from which factors are extracted** | **cov**ariance | Matrix of corrs | Can only be used with pca; preceded by specification of number of factors | | pca a b c, cov<br>pca a b c, fa(3) cov<br><br>pca a b c, pf mine(1) cov |
| **Plot eigenvalues** | **screeplot** | | Running of factor command | | screeplot |

### (1.2) for correlation matrices

You may want to perform a factor analysis directly from a correlation matrix rather then form variables in a dataset. Since you may not have access to the dataset, or you may have used another method of estimating a correlation matrix-e.g., as amatrix of tetrachoric correlation (as part(5) ). See help for more detail.

| Description | Command | Preceded by | Followed by | Example: |
|---|---|---|---|---|
| **Principal components analysis** | **pcamat** *matname ,n(#)* | | Correlation matrix | **pcamat** *P, n(789)* |

## (2)Factor Analysis

### (2.1) for variables

| Description | Command | | default | Preceded by | Followed by | Example: |
|---|---|---|---|---|---|---|
| **Performs factor analysis** | **Fac**tor | | | | | fac a b c |
| **Type of estimation (extraction method)** | **pf** | principal factor | pf | | | fac a b c, ml |
| | **pcf** | Principle components factor | | | | |
| | **ipf** | iterated principal factor | | | | |
| | **ml** | Maximum likelihood | | | | |
| **Specifies number of factors** | **fa**ctors(as an option) | | # of vars | Extraction method | The number of desired factors in parentheses | fac a b c, pf fa(3) |
| **Keeps factors with eigenvalues > #specified** | **mine**igen | | Default for all methods except pcf | Extraction method | The cutoff eigenvalue in parentheses | fac a b c, mine(1) |

| | | | | | |
|---|---|---|---|---|---|
| | | is 0 ;The default for pcf is 1 | | | |
| **Display loadings as blank when \|loadings\|<#** | **blanks(#)** | | Extraction method | | fac a b c, fa(2) blanks(0.30) |

**quick note**: in commands, for example *factor*, you can type the whole word, or just the first few letters **in bold , fac**.

**(2.2) for correlation matrix**

| Description | Command | Preceded by | Followed by | Example: |
|---|---|---|---|---|
| **Factor analysis** | **factormat** *matname ,n(#)* | | Correlation matrix | *factormat P,n(979) factors(2) ipf* |

## (3) factor analysis post-estimation commands

| Description | Command | default | Preceded by | Example: |
|---|---|---|---|---|
| **Rotates extracted factors: Varimax or oblique rotation** | **rot**ate | Varimax | Running of factor command | rot |
| **Specifies (oblique) rotation** | **p**romax | Promax(3) | Factor command | rot, p(2)<br>rot, p |
| | **oblimin oblique** | | Factor command | rot, oblimin oblique |
| **Plot factor loadings** | **loadingplt** | | Factor command | loadingplot |
| **Predict regression or Bartlett scores** | **predict** | Regression scores | Running of factor command or rotation | predict u1 u2 u3, bartlett |
| **Gives you scores from unrotated matrix** | **norotate** | If you have already rotated, and you don't specify *norotate* it will give you scores from the rotated factors | Predict command and list of factor names and a comma | predict u1 u2 u3, norotate |

**Another Quick Notes:**
As a time saver, if you varlist consisted of variables with the same prefix followed by consecututive numbers (i.e, *slf961,slf962,slf963*) then you could type the factor command this way: fac slf961-slf965

Now, even though after you run the factor command it shows you all of the eigenvalues, it will only keep the ones you specify for use through either factors(#) or mineigen(n). Those eigenvalues and corresponding eigenvectors are then saved for use with either the **predict** or **rot**ate commands.

**Now something odd about stata:** when you do the score command, you are, perhaps unbeknownst, naming the factors that have just been extracted. If you wanted to get factor scores before and after rotation, then you would give the unrotated factors one set of names, and the rotated factors another set of names. As soon as you name the factors, if you look at your dataset, you will see that those variables have been added at the end of your data.

## (5)The following postestimation commands are of special interests after factor and factormat :

- **estat common** reports the correlation matrix of the common factors and is more of interest after oblique rotations.
- **estat factors** reports model-selection criteria (AIC and BIC) over all the factors retained in an analysis.
- **estat rotatecompare** reports the unrotated factor loadings next to the most-recent rotated loadings.
- **estat structure** reports the factor structure — the correlations between the variables and the common factors.
- **estat rotatecompare** reports the unrotated (principal) components next to the most recent rotated components
- **estat anti** reports the anti-image correlation and anti-image covariance matrices.
- **estat kmo** reports the Kaiser–Meyer–Olkin measure of sampling adequacy.
- **estat residuals** reports the difference between the observed correlation or covariance matrix and the fitted (reproduced) matrix using the retained factors.
- **estat smc** reports the squared multiple correlations (SMC) between each variable and all other variables. SMC is a theoretical lower bound for communality, so it is an upper bound for the unexplained variance.

***More detail about factor analysis , please refer to "STATA MULTIVARIATE STATISTCIS REFERENCE MANUAL RELEASE 9" or search for help command .***

## (5)Sample Example

. factor slf941 slf942 slf943 slf944, fa(2)
(obs=328)

```
            (principal factors; 1 factor retained)
Factor      Eigenvalue     Difference    Proportion    Cumulative
-------------------------------------------------------------------
     1        1.77520         1.83445      1.2556         1.2556
     2       -0.05925         0.06636     -0.0419         1.2137
     3       -0.12561         0.05094     -0.0888         1.1249
     4       -0.17656               .     -0.1249         1.0000

            Factor Loadings
 Variable |      1    Uniqueness
----------+--------------------
   slf941 |   0.66556    0.55703
   slf942 |   0.73133    0.46516
   slf943 |   0.60900    0.62912
   slf944 |   0.65308    0.57349
```

. rot

```
            (varimax rotation)
            Rotated Factor Loadings
 Variable |      1    Uniqueness
```

```
----------+---------------------
   slf941 |   0.66556    0.55703
   slf942 |   0.73133    0.46516
   slf943 |   0.60900    0.62912
   slf944 |   0.65308    0.57349
```

## (6)Syntax for logistic and ROC curve

**Syntax:**

logistic *depvar varlist* /***logit also works***/
lroc /***To make ROC curve***/
roctab *depvar varlist*, detail /* **displaying the sensitivity, specificity, calculates the area under the ROC curve** */

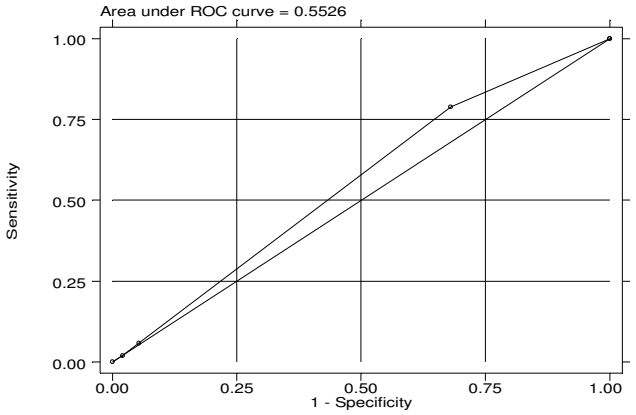
**example:**

logistic mari1594 slf941

Logit estimates Number of obs = 296
LR chi2(1) = 1.42
Prob > chi2 = 0.2331
Log likelihood = -136.86173 Pseudo R2 = 0.0052

```
------------------------------------------------------------------------------
mari1594 | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  slf941 |   .7445419   .1826371    -1.203   0.229      .4603501    1.204176
------------------------------------------------------------------------------
```

lroc




```
. roctab mari1594 slf941, detail
Detailed report of Sensitivity and Specificity
------------------------------------------------------------------------------
                                          Correctly
Cutpoint      Sensitivity   Specificity   Classified          LR+          LR-
------------------------------------------------------------------------------
( >= 1 )            0.00%       100.00%       82.43%                    1.0000
( >= 2 )           21.15%        68.03%       59.80%       0.6617       1.1589
( >= 3 )           94.23%         5.33%       20.95%       0.9953       1.0828
( >= 4 )           98.08%         2.05%       18.92%       1.0013       0.9385
( >  4 )          100.00%         0.00%       17.57%       1.0000
------------------------------------------------------------------------------
```