Generating normal variables

The following PRELIS input file (**SIMEX11.PRL** in the **PRELIS Examples** folder) generates 200 independent cases of six variables having a multivariate normal distribution with zero mean vector and covariance matrix $\boldsymbol{\Sigma}$. Note that there is no NI value specified on the DA command.

```
Generating Multivariate Normal Variables
with a Specified Covariance Matrix
DA NO=200
NE V1=NRAND
NE V2=NRAND
NE V3=NRAND
NE V4=NRAND
NE V5=NRAND
NE V6=NRAND
NE X1=V1
NE X2=.378*V1+.925806*V2
NE X3=.72*V1+.068956*V2+.690540*V3
NE X4=.324*V1+.321372*V2+.047151*V3+.88855*V4
NE X5=.27*V1+.26781*V2+.039292*V3+.140229*V4+.913329*V5
NE X6=.27*V1+.025858*V2+.063453*V3+.010374*V4+.006818*V5+.960339*V6
CO ALL
SD V1-V6
OU RA=RAWDATA WI=7 ND=3 XM IX=123456
```

Here we generate data on V1 through V6 which are independent and normally distributed with mean zero and variance one. V1 through V6 are then transformed linearly to X1 through X6 using the $\mathbf{T}$ matrix. After X1 through X6 have been defined, V1 through V6 can be deleted. This is done with the SD command.

The raw data obtained in the file **RAWDATA** in the format 6F7.3. WI = 7 defines the width of each field and ND = 3 defines the number of decimals for the data in the file **RAWDATA**.

Now suppose we are not interested in the sample of raw data but only in the sample covariance matrix. Then, just replace the **RA = RAWDATA** specification with **CM = SIMEX1.CM**, see file **SIMEX12.PRL**. Then no raw data will be saved or stored. The covariance matrix will be computed "on the run".

Next, suppose we want to generate 400 samples of size 200 and save all the 400 sample covariance matrices. Then, just add RP = 400 on the DA command, see file **SIMEX13.PRL**. The file **SIMEX1.CM** will then contain 400 sample covariance matrices. The first two replicates in **SIMEX1.CM** are:

```
 0.95375D+00  0.37841D+00  0.10018D+01  0.59348D+00  0.30790D+00  0.83902D+00
 0.36562D+00  0.40900D+00  0.32791D+00  0.11061D+01  0.20653D+00  0.30234D+00
 0.19845D+00  0.28648D+00  0.87716D+00  0.22175D+00 -0.78280D-02  0.14026D+00
-0.12499D-01  0.32627D-01  0.10423D+01
 0.87794D+00  0.37148D+00  0.12746D+01  0.60429D+00  0.41321D+00  0.86589D+00
 0.29965D+00  0.49782D+00  0.28989D+00  0.92703D+00  0.14368D+00  0.39918D+00
 0.15315D+00  0.26849D+00  0.98916D+00  0.30765D+00  0.17819D+00  0.25911D+00
 0.30325D-01  0.40118D-01  0.11179D+01
```

This file can be read by LISREL to estimate a model for each sample. The following input file (**SIMEX14.LIS**) will estimate a confirmatory factor analysis model with two correlated factors, see Jöreskog (1979), Hägglund (1982), or Jöreskog & Sörbom (1989b).

```
Fitting a confirmatory factor analysis model in each of 400 samples
DA NI=6 NO=200 RP=400
CM=SIMEX1.CM
MO NX=6 NK=2
FR LX(1,1) LX(2,2) LX(3,1) LX(4,2) LX(5,2) LX(6,1)
OU AD=OFF PV=SIMEX1.PV SV=SIMEX1.SV GF=SIMEX1.GF XM
```

There are 13 free parameters in the model. The file **SIMEX1.PV** contains 400 sets of 13 estimated parameters, one from each sample. The file **SIMEX1.SV** contains 400 sets of 13 estimated standard errors of these parameter estimates, one from each sample. The file **SIMEX1.GF** contains 400 sets of all the 42 fit measures that LISREL computes for each sample. The information in each of these files may be studied further. Here we show how the parameter estimates in **SIMEX1.PV** may be examined.

The first two sets of parameter estimates in **SIMEX1.PV** are:

```
     1  0  0
  0.853458D+00     0.660581D+00     0.698258D+00     0.641822D+00     0.433952D+00
0.224460D+00
  0.655037D+00     0.225359D+00     0.565433D+00     0.351457D+00     0.694165D+00
0.688845D+00
  0.991918D+00
     2  0  0
  0.781581D+00     0.863154D+00     0.774720D+00     0.592991D+00     0.438305D+00
0.357914D+00
  0.578626D+00     0.267072D+00     0.529566D+00     0.265699D+00     0.575391D+00
0.797048D+00
  0.989797D+00
```

The first number is the replication number, the second is an error indicator which is

- 0 if iterations have converged and the *P*-value for χ^2 is in the interval $0.0005 \le P \le 0.9995.$
- 1 if iterations have not converged
- 2 if iterations have converged and the *P*-value for χ^2 is either $P < 0.0005$ or $P > 0.9995$. In this case, confidence limits for the fit statistics in file **SIMEX1.GF** have not been computed.

The following PRELIS input file (**SIMEX15.PRL**) analyzes the parameter estimates in the file **SIMEX1.PV**.

```
Analyzing the parameter estimates in SIMEX1.PV
DA NI=14
LA
IND 'LX(1,1)' 'LX(2,2)' 'LX(3,1)' 'LX(4,2)' 'LX(5,2)' 'LX(6,1)'
         'PH(1,1)' 'TD(1)' 'TD(2)' 'TD(3)' 'TD(4)' 'TD(5)' 'TD(6)'
RA=SIMEX1.PV FO;(3X,F3.0/(6D14.6))
CO 'LX(1,1)' - 'TD(6)'
OU
```

The output file gives the following information about the sampling distribution of the parameter estimates.

```
Univariate Summary Statistics for Continuous Variables

 Variable     Mean  St. Dev.  Skewness  Kurtosis  Minimum Freq.  Maximum Freq.
 --------     ----  --------  --------  --------  ------- -----  ------- -----
      IND  200.500   115.614     0.000    -1.200    1.000     1  400.000     1
  LX(1,1)    0.903     0.076     0.044    -0.305    0.705     1    1.136     1
  LX(2,2)    0.697     0.085     0.115     0.091    0.485     1    0.956     1
  LX(3,1)    0.801     0.074    -0.159    -0.046    0.560     1    0.990     1
  LX(4,2)    0.602     0.079    -0.186     0.256    0.312     1    0.839     1
  LX(5,2)    0.499     0.080    -0.141    -0.024    0.233     1    0.715     1
  LX(6,1)    0.299     0.072    -0.066     0.059    0.089     1    0.489     1
  PH(1,1)    0.606     0.077    -0.346     0.248    0.327     1    0.796     1
    TD(1)    0.181     0.091    -0.315    -0.092   -0.075     1    0.401     1
    TD(2)    0.511     0.097    -0.175     0.560    0.151     1    0.814     1
    TD(3)    0.356     0.078    -0.208     0.726    0.053     1    0.577     1
    TD(4)    0.627     0.087     0.115     0.070    0.382     1    0.902     1
    TD(5)    0.742     0.087     0.115    -0.098    0.437     1    0.968     1
    TD(6)    0.901     0.095     0.419     0.473    0.667     1    1.308     1

 Test of Univariate Normality for Continuous Variables

              Skewness          Kurtosis       Skewness and Kurtosis

 Variable Z-Score P-Value   Z-Score P-Value   Chi-Square P-Value

      IND   0.000   1.000   -15.408   0.000      237.417   0.000
  LX(1,1)   0.366   0.714    -1.385   0.166        2.054   0.358
  LX(2,2)   0.949   0.343     0.479   0.632        1.129   0.569
```

LX(3,1)	-1.308	0.191	-0.086	0.932	1.717	0.424
LX(4,2)	-1.526	0.127	1.073	0.283	3.478	0.176
LX(5,2)	-1.159	0.247	0.010	0.992	1.342	0.511
LX(6,1)	-0.546	0.585	0.355	0.723	0.424	0.809
PH(1,1)	-2.793	0.005	1.046	0.295	8.897	0.012
TD(1)	-2.550	0.011	-0.292	0.770	6.587	0.037
TD(2)	-1.437	0.151	1.983	0.047	5.998	0.050
TD(3)	-1.704	0.088	2.406	0.016	8.689	0.013
TD(4)	0.948	0.343	0.396	0.692	1.056	0.590
TD(5)	0.953	0.340	-0.315	0.752	1.008	0.604
TD(6)	3.338	0.001	1.741	0.082	14.175	0.001

This reveals that:

- The mean of sampling distribution is close to the true population value for most parameters.
- There are some negative parameter estimates of the error variance TD(1) (Heywood cases), see comment in Jöreskog & Sörbom (1989).
- The asymptotic normality approximation, which LISREL assumes when estimating standard errors of parameter estimates, may not be sufficiently close for some of the parameters.

This suggest that the sample size should be larger than 200 to avoid Heywood cases and to obtain correct standard errors.