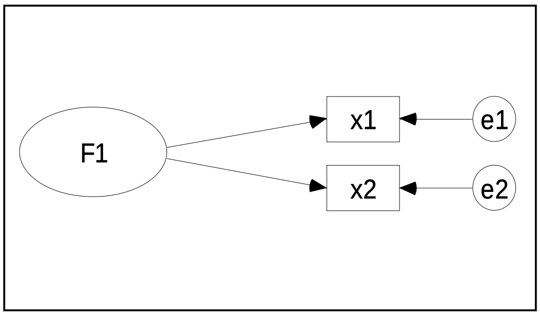
共分散構造分析プログラムによる主成分分析

2000.1.15

観測変数が２個の場合を考える．因子分析のモデル式表現とパス図表現は,

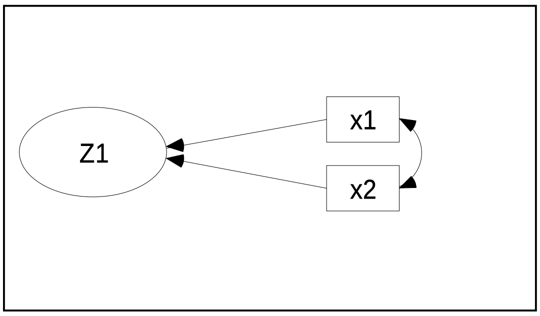
$$x1 = \lambda_1 F1 + e1$$
$$x2 = \lambda_2 F1 + e2$$



主成分分析のモデル式表現は.

$$Z_1 = a_1\, x1 + a_2\, x2$$

だから，そのパス図表現は,



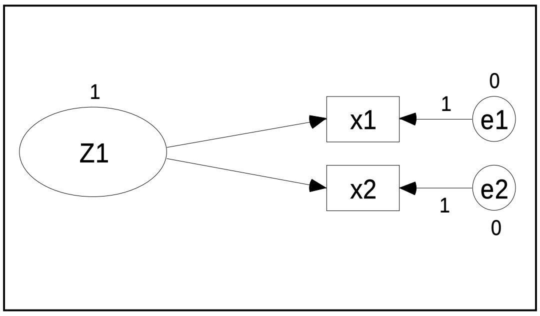
で良さそうな気がする.
Amos のようにパス図から共分散構造分析を実行できるソストウエアで実際に上図のようなパス図を描いてもエラーとなって解を求めることはできない．共分散構造分析モデルの枠組みでは上図には従属変数 Z1 に誤差変数を要求し，仮に誤差分散を０に固定しても自由度が負となってモデルを識別できない.

主成分分析と因子分析は似ているといわれるのだが，観測変数の線形結合である主成分分析をあえて因子分析モデル（共分散構造モデル）の枠組みで表現するならば，以下のような方程式表現となる.

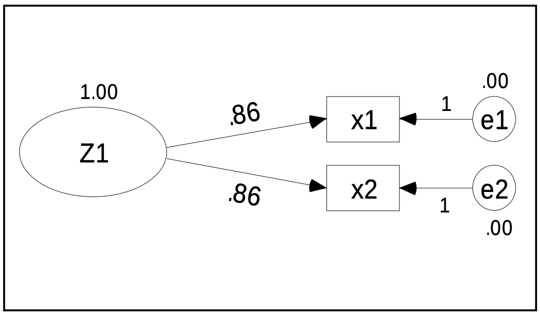
$$x1 = \lambda_1 F1 + 0$$
$$x2 = \lambda_2 F1 + 0$$

つまり誤差項のない（=0）ような因子分析モデルとなる.

このモデルは観測変数と同じ個数まで因子を抽出すれば正確に相関行列を分解できるけれど，次元縮小という多変量解析の目標からみれば，それは単なる数学的変換に過ぎない．そこで因子分析の主成分解とは，最初の少数個の成分だけを採用し，残りをひとまとめに独自因子とみなして因子分析モデルを近似しようとする方法であった．主成分の分散を固有値ではなく１とし，独自因子に直交仮定を置かないモデルが因子分析の主成分解である．主成分分析（因子分析の主成分解）をパス図で描くと以下のようになる[1].



主成分（因子）Z1 の分散を１とし．誤差分散を０に固定している．このモデルを最小２乗法で第 1 主成分を求めると，以下の推定値が得られる．



パス係数に表示されているのは因子負荷量であり固有ベクトルの値ではない．ここで使用した数値例の相関行列は２変数の相関係数が 0.5 の場合である．

$$\begin{bmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{bmatrix}$$

２変数だけの相関行列の主成分は，奥野他(1981)『多変量解析法《改定版》』にあるように，幾何学的には座標を 45 ° 回転した結果であり，固有値 λ は[2]相関係数 r との関係で，

$$\lambda_1 = 1 + |r| = 1.5$$
$$\lambda_2 = 1 - |r| = 0.5$$

となる．固有ベクトルの２つの要素は等しい値で，相関係数と関係なく常に，

$$a_1 = 1/\sqrt{2} = 0.707$$

[1] 主成分分析を SEM で表現する詳細は豊田（1992）SAS による共分散構造分析（東大出版）を参照.
[2] ここの記号λは固有値として使用．先の文脈で因子負荷として使用したλとは別.

従って，固有値の平方根を乗じた因子負荷量（構造係数）も等しくて常に.

$$l_1 = \quad 0.707 * \sqrt{1.5} = \quad 0.866$$

となる．これが先のパス図に示されたパス係数（因子負荷）である．このパス図は Amos4 による結果で，小数点以下第 2 位に誤差があって「0.86」と表示されているが，正しくは「0.87」である．

参考までに SAS による主成分分析（因子分析の主成分解）のプログラムと出力を掲載する．

```
data x( type = corr ) ;
   input _type_$ _name_$ x1 x2 ;
cards ;
N     .  100 100
MEAN .  0.0 0.0
CORR x1 1.0 0.5
CORR x2 0.5 1.0
;
run ;
proc factor data = x ;
run ;
```

```
                         Prior Communality Estimates: ONE

            Eigenvalues of the Correlation Matrix:  Total = 2  Average = 1

                                          1           2
                         Eigenvalue      1.5000      0.5000
                         Difference      1.0000
                         Proportion      0.7500      0.2500
                         Cumulative      0.7500      1.0000

                  1 factors will be retained by the MINEIGEN criterion.

                                  Factor Pattern

                                            FACTOR1

                                  X1        0.86603
                                  X2        0.86603
```

SAS の CALIS プロシジャの lineqs ステートメント（EQS モデル）の書法と出力結果は以下のようになる．

```
proc calis data = x method = uls  ;
    var x1 x2 ;
    lineqs
        x1 = a1 f1 + e1 ,
        x2 = a2 f1 + e2 ;
    std f1 = 1 , e1 = 0 , e2 = 0 ;
run ;
```

```
           Covariance Structure Analysis: Least-Squares Estimation

                        Levenberg-Marquardt Optimization
                         Scaling Update of More (1978)
                        Number of Parameter Estimates 2
                      Number of Functions (Observations) 3

Optimization Start: Active Constraints= 0  Criterion= 0.878
Maximum Gradient Element= 0.923 Radius= 1.164

         Iter rest nfun act   optcrit  difcrit maxgrad  lambda     rho
           1    0    3   0    0.1668   0.7113   0.473   1.347   1.280
           2    0    4   0    0.1252   0.0416  0.0406       0   1.002
           3    0    5   0    0.1250 0.000196 0.00382       0   1.075
           4    0    6   0    0.1250 6.302E-6 0.00126       0   1.333
           5    0    7   0    0.1250 7.002E-7 0.00042       0   1.333
           6    0    8   0    0.1250  7.78E-8 0.00014       0   1.333
           7    0    9   0    0.1250 8.645E-9 0.00005       0   1.333
           8    0   10   0    0.1250 9.61E-10 0.00002       0   1.333
           9    0   11   0    0.1250 1.07E-10 5.17E-6       0   1.333

Optimization Results: Iterations= 9 Function Calls= 12 Jacobian Calls= 10
Active Constraints= 0  Criterion= 0.125 Maximum Gradient Element= 5.16539E-6
Lambda= 0 Rho= 1.333 Radius= 0.00003116

NOTE: GCONV convergence criterion satisfied.

           Covariance Structure Analysis: Least-Squares Estimation

          Fit criterion . . . . . . . . . . . . . . . . . .     0.1250
          Goodness of Fit Index (GFI) . . . . . . . . . . .     0.9000
          GFI Adjusted for Degrees of Freedom (AGFI). . . .     0.7000
          Root Mean Square Residual (RMR) . . . . . . . . .     0.2500
          Parsimonious GFI (Mulaik, 1989) . . . . . . . . .     0.9000

           Covariance Structure Analysis: Least-Squares Estimation

                         Manifest Variable Equations

                      X1       =     0.8660*F1 + 1.0000 E1
                                            A1

                      X2       =     0.8660*F1 + 1.0000 E2
                                            A2

                      Variances of Exogenous Variables
                    -------------------------------------
                    Variable    Parameter       Estimate
                    -------------------------------------
                    F1                          1.000000

           Covariance Structure Analysis: Least-Squares Estimation

                   Equations with Standardized Coefficients

                      X1       =     1.0000*F1 + 0.0000 E1
                                            A1

                      X2       =     1.0000*F1 + 0.0000 E2
                                            A2

                         Squared Multiple Correlations
          ----------------------------------------------------------
                             Error            Total
           Variable        Variance         Variance         R-squared
          ----------------------------------------------------------
             1     X1             0         0.749996          1.000000
             2     X2             0         0.750004          1.000000
```