Path Analysis

Factor analysis is a method used to identify "latent" variables which theoretically may underlie observed variables in common with those latent variables or factors. For example, a factor analysis of several intelligence, aptitude and achievement test batteries may suggest only two or three underlying factors can "explain" a large proportion of common variance among the many subtests of the original measures. Two kinds of factor analytic uses are present in the literature: exploratory and confirmatory. In exploratory FA, the researcher is attempting to develop plausible hypotheses about possible underlying common variables. In confirmatory FA, the research is typically confirming the findings of prior research or theory. For both applications, there are a variety of alternative approaches based on, for example, whether or not one is interested in estimating results for a population of variables or a population of subjects. The researcher must also decide if his or her theory "makes sense" to have correlated factors or uncorrelated (orthogonal) factors. Because FA can result in an infinite number of possible solutions as a function of how the axis of the factors are rotated and correlated, the researcher must decide on the method of rotation which is most likely to lead to a simple yet plausible solution (parsimoneous solution.)

In the Factor Analysis dialog you select the variables from the left list and enter them by clicking those selected and clicking the left arrow. Click on the Output Options desired, the method of analysis desired and the method of rotation desired. (Not all rotation options may currently be implemented in your version.) The Varimax method is frequently employed and a good place to begin. For a confirmatory analysis, you will likely select the maximum number of factors to correspond to that of the theory or previous findings. The Procrustean rotation method would also be selected in a confirmatory analysis. In that option, you will be asked to enter the factor loadings of a theoretical (or previous solution) as the "target" matrix with which to rotate the current solution to maximum congruity.

Example of a Path Analysis

In this example we will use the file CANSAS.LAZ. The user begins by selecting the Path Analysis option of the Statistics / Multivariate menu. In the figure below we have selected all variables to analyze and have entered our first path indicating that waist size is "caused" by weight:

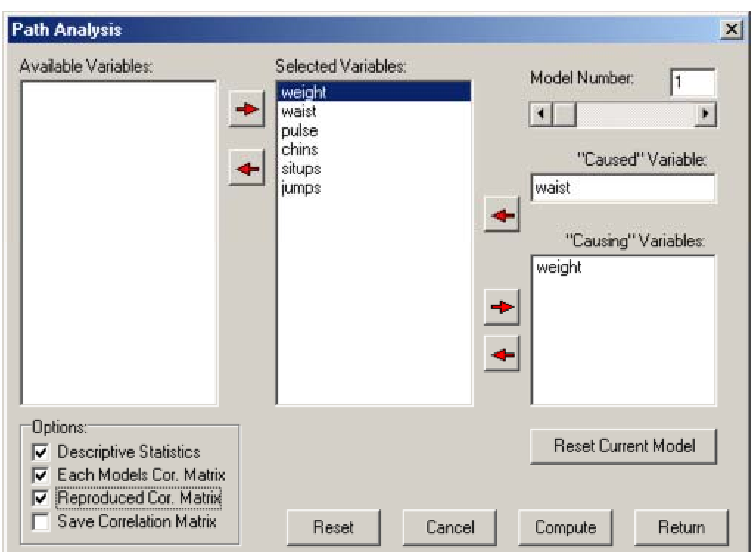

Figure 1. Path Analysis Form

We will also hypothesize that pulse rate is “caused” by weight, chin-ups are “caused” by weight, waist and pulse, that the number of sit-ups is “caused” by weight, waist and pulse and that jumps are “caused” by weight, waist and pulse. Each time we enter a new causal relationship we click the scroll bar to move to a new model number prior to entering the “caused” and “causing” variables. Once we have entered each model, we then click on the Compute button. Note we have elected to print descriptive statistics, each models correlation matrix, and the reproduced correlation matrix which will be our measure of how well the models “fit” the data. The results are shown below:

```
PATH ANALYSIS RESULTS

CAUSED VARIABLE: waist
     Causing Variables:
    weight
CAUSED VARIABLE: pulse
     Causing Variables:
    weight
CAUSED VARIABLE: chins
     Causing Variables:
    weight
    waist
    pulse
CAUSED VARIABLE: situps
     Causing Variables:
    weight
    waist
    pulse
CAUSED VARIABLE: jumps
     Causing Variables:
    weight
    waist
    pulse

Correlation Matrix with   20 valid cases.

Variables
                weight        waist        pulse        chins       situps
   weight        1.000        0.870       -0.366       -0.390       -0.493
    waist        0.870        1.000       -0.353       -0.552       -0.646
    pulse       -0.366       -0.353        1.000        0.151        0.225
    chins       -0.390       -0.552        0.151        1.000        0.696
   situps       -0.493       -0.646        0.225        0.696        1.000
     jumps      -0.226       -0.191        0.035        0.496        0.669


Variables
                  jumps
   weight       -0.226
    waist       -0.191
    pulse        0.035
    chins        0.496
   situps        0.669
     jumps       1.000


MEANS with   20 valid cases.

Variables      weight        waist        pulse        chins       situps
              178.600       35.400       56.100        9.450      145.550

Variables        jumps
                70.300

VARIANCES with   20 valid cases.

Variables      weight        waist        pulse        chins       situps
              609.621       10.253       51.989       27.945     3914.576

Variables        jumps
              2629.379
```

```
STANDARD DEVIATIONS with   20 valid cases.

Variables      weight        waist        pulse        chins       situps
                24.691        3.202        7.210        5.286       62.567

Variables        jumps
                51.277

Dependent Variable = waist

Correlation Matrix with   20 valid cases.

Variables
                weight        waist
   weight        1.000        0.870
    waist        0.870        1.000


MEANS with   20 valid cases.

Variables      weight        waist
               178.600       35.400


VARIANCES with   20 valid cases.

Variables      weight        waist
               609.621       10.253


STANDARD DEVIATIONS with   20 valid cases.

Variables      weight        waist
                24.691        3.202

Dependent Variable = waist

       R        R2         F     Prob.>F  DF1  DF2
   0.870     0.757    56.173     0.000    1   18
Adjusted R Squared = 0.744

Std. Error of Estimate =      1.621

Variable       Beta      B         Std.Error t         Prob.>t
   weight      0.870     0.113     0.015     7.495     0.000

Constant =     15.244

Dependent Variable = pulse

Correlation Matrix with   20 valid cases.
Variables
                weight        pulse
   weight        1.000       -0.366
    pulse       -0.366        1.000


MEANS with   20 valid cases.
Variables      weight        pulse
               178.600       56.100


VARIANCES with   20 valid cases.
Variables      weight        pulse
               609.621       51.989
```

STANDARD DEVIATIONS with 20 valid cases.

Variables	weight	pulse
	24.691	7.210

Dependent Variable = pulse

R	R2	F	Prob.>F	DF1	DF2
0.366	0.134	2.780	0.113	1	18

Adjusted R Squared = 0.086

Std. Error of Estimate = 6.895

Variable	Beta	B	Std.Error	t	Prob.>t
weight	-0.366	-0.107	0.064	-1.667	0.113

Constant = 75.177

Dependent Variable = chins

Correlation Matrix with 20 valid cases.

Variables	weight	waist	pulse	chins
weight	1.000	0.870	-0.366	-0.390
waist	0.870	1.000	-0.353	-0.552
pulse	-0.366	-0.353	1.000	0.151
chins	-0.390	-0.552	0.151	1.000

MEANS with 20 valid cases.

Variables	weight	waist	pulse	chins
	178.600	35.400	56.100	9.450

VARIANCES with 20 valid cases.

Variables	weight	waist	pulse	chins
	609.621	10.253	51.989	27.945

STANDARD DEVIATIONS with 20 valid cases.

Variables	weight	waist	pulse	chins
	24.691	3.202	7.210	5.286

Dependent Variable = chins

R	R2	F	Prob.>F	DF1	DF2
0.583	0.340	2.742	0.077	3	16

Adjusted R Squared = 0.216

Std. Error of Estimate = 4.681

Variable	Beta	B	Std.Error	t	Prob.>t
weight	0.368	0.079	0.089	0.886	0.389
waist	-0.882	-1.456	0.683	-2.132	0.049
pulse	-0.026	-0.019	0.160	-0.118	0.907

Constant = 47.968

Dependent Variable = situps

Correlation Matrix with 20 valid cases.

Variables	weight	waist	pulse	situps
weight	1.000	0.870	-0.366	-0.493
waist	0.870	1.000	-0.353	-0.646
pulse	-0.366	-0.353	1.000	0.225
situps	-0.493	-0.646	0.225	1.000

MEANS with 20 valid cases.

Variables	weight	waist	pulse	situps
	178.600	35.400	56.100	145.550

VARIANCES with 20 valid cases.

Variables	weight	waist	pulse	situps
	609.621	10.253	51.989	3914.576

STANDARD DEVIATIONS with 20 valid cases.

Variables	weight	waist	pulse	situps
	24.691	3.202	7.210	62.567

Dependent Variable = situps

R	R2	F	Prob.>F	DF1	DF2
0.661	0.436	4.131	0.024	3	16

Adjusted R Squared = 0.331

Std. Error of Estimate = 51.181

Variable	Beta	B	Std.Error	t	Prob.>t
weight	0.287	0.728	0.973	0.748	0.466
waist	-0.890	-17.387	7.465	-2.329	0.033
pulse	0.016	0.139	1.755	0.079	0.938

Constant = 623.282

Dependent Variable = jumps

Correlation Matrix with 20 valid cases.
Variables

	weight	waist	pulse	jumps
weight	1.000	0.870	-0.366	-0.226
waist	0.870	1.000	-0.353	-0.191
pulse	-0.366	-0.353	1.000	0.035
jumps	-0.226	-0.191	0.035	1.000

MEANS with 20 valid cases.

Variables	weight	waist	pulse	jumps
	178.600	35.400	56.100	70.300

VARIANCES with 20 valid cases.

Variables	weight	waist	pulse	jumps
	609.621	10.253	51.989	2629.379

STANDARD DEVIATIONS with 20 valid cases.

Variables	weight	waist	pulse	jumps
	24.691	3.202	7.210	51.277

Dependent Variable = jumps

R	R2	F	Prob.>F	DF1	DF2
0.232	0.054	0.304	0.822	3	16

Adjusted R Squared = -0.123

Std. Error of Estimate = 54.351

Variable	Beta	B	Std.Error	t	Prob.>t
weight	-0.259	-0.538	1.034	-0.520	0.610
waist	0.015	0.234	7.928	0.029	0.977
pulse	-0.055	-0.389	1.863	-0.209	0.837

Constant = 179.887

Matrix of Path Coefficients with 20 valid cases.
Variables

	weight	waist	pulse	chins	situps
weight	0.000	0.870	-0.366	0.368	0.287
waist	0.870	0.000	0.000	-0.882	-0.890
pulse	-0.366	0.000	0.000	-0.026	0.016
chins	0.368	-0.882	-0.026	0.000	0.000
situps	0.287	-0.890	0.016	0.000	0.000
jumps	-0.259	0.015	-0.055	0.000	0.000

```
Variables
                  jumps
   weight       -0.259
    waist        0.015
    pulse       -0.055
    chins        0.000
   situps        0.000
     jumps       0.000

SUMMARY OF CAUSAL MODELS
Var. Caused    Causing Var.  Path Coefficient
      waist         weight    0.870
      pulse         weight   -0.366
      chins         weight    0.368
      chins          waist   -0.882
      chins          pulse   -0.026
     situps         weight    0.287
     situps          waist   -0.890
     situps          pulse    0.016
       jumps        weight   -0.259
       jumps         waist    0.015
       jumps         pulse   -0.055

Reproduced Correlation Matrix with   20 valid cases.
Variables
                weight        waist        pulse        chins       situps
   weight        1.000        0.870       -0.366       -0.390       -0.493
    waist        0.870        1.000       -0.318       -0.553       -0.645
    pulse       -0.366       -0.318        1.000        0.120        0.194
    chins       -0.390       -0.553        0.120        1.000        0.382
   situps       -0.493       -0.645        0.194        0.382        1.000
     jumps      -0.226       -0.193        0.035        0.086        0.108


Variables
                  jumps
   weight       -0.226
    waist       -0.193
    pulse        0.035
    chins        0.086
   situps        0.108
     jumps       1.000

Average absolute difference between observed and reproduced
coefficients := 0.077
Maximum difference found := 0.562
```

We note that pulse is not a particularly important predictor of chin-ups or sit-ups. The largest discrepancy of 0.562 between an original correlation and a correlation reproduced using the path coefficients indicates our model of causation may have been inadequate.