

Path Analysis Using PROC CALIS

SAS Code:

```

/* file ~carey/p7291dir/pathreg2.sas
   Example of using PROC CALIS to perform multivariate
   multiple regression. The PROC REG commands are on
   file ~carey/pathreg.sas */

OPTIONS NOCENTER NODATE;
TITLE1 'Path Analysis on the Interest Data set using PROC CALIS';

/* The COV option to PROC CALIS instructs CALIS to analyze the
   covariance matrix instead of the correlation matrix
   (the default).
   This will give us unstandardized as well as standardized
   results.
   The CORR option prints the covariance and correlation matrix.
   The RESIDUAL option prints the difference between the observed
   covariance matrix and the predicted covariance matrix.
*/
PROC CALIS DATA=mvstats.interest COV CORR RESIDUAL;

/* the VAR statement gives the observed variables to be analyzed */
VAR lawyer archtct educ vocab geometry;

/* LINEQS gives the LINEar EQUationS. There should be a separate
   equation for each endogenous variable. Each equation should be
   separated by a comma and a semicolon should be placed after the
   last equation. The equations are all of the form:

           Y = b X + b X + ... b X + e
           i   i1 1   i2 2       ip P   i

The Ys and the Xs should be names of variables. The bs can be
any name that you wish as long as it is not a variable name.
The error term MUST start with an e. */
LINEQS
    lawyer = ble educ + blv vocab + blg geometry + e_l,
    archtct = bae educ + bav vocab + bag geometry + e_a;

/* The STD statement gives names for the variances (not the
   standard deviations) of the exogenous variables */
STD educ=v_edu, vocab=v_voc, geometry=v_geo, e_l=ve_l, e_a=ve_a;

/* The COV statement gives names for the covariances between
   the exogenous variables */
COV educ vocab =cov_ev, educ geometry=cov_eg,
    vocab geometry = cov_vg, e_l e_a = c_elea ;

RUN;

```

OUTPUT:

Path Analysis on the Interest Data set using PROC CALIS
Covariance Structure Analysis: Pattern and Initial Values

The output below gives the path model in terms of the matrix notation used by Peter Bentler's EQS program. For this example, only two matrices are used: the `_PHI_` matrix (the covariance matrix for the exogenous variables) and the `_GAMMA_` matrix (the path coefficients between all 7 variables in the model and the 5 exogenous variables). The `_BETA_` matrix is for the path coefficients among the endogenous variables.

```

LINEQS Model Statement
-----
      Matrix          Rows & Cols          Matrix Type
TERM  1-----
  1   _SEL_           5         7   SELECTION
  2   _BETA_          7         7   EQSBETA      IMINUSINV
  3   _GAMMA_         7         5   EQSGAMMA
  4   _PHI_           5         5   SYMMETRIC

```

There are several important definitions for the variables used by CALIS when the LINEQS statement is invoked.

Exogenous variables are those that do not have a single-headed arrow entering them in a path diagram and can appear only on the right hand side of a structural equation.

Endogenous variables are those that have at least one single-headed arrow entering them and appear on the left hand side of a structural equation. They may also appear on the right hand side of the equation in some models.

Manifest variables are variables on which the observations have concrete numbers. Manifest variables can be either exogenous or endogenous.

Latent variables are variables on which the observations have no concrete numbers. They may be either endogenous or exogenous.

Error variables are residuals for manifest endogenous variables. Each manifest endogenous variable in a model should have an error variable associated with it. Error variables are always exogenous. In the LINEQS statement, error variables must begin with the letter "e".

Disturbance variables are residuals for latent endogenous variables. Each latent endogenous variable should have a disturbance associated with it. In the LINEQS statement, disturbance variables must begin with the letter "d".

The following output gives the variables and their type in the current model. Note that there are latent variables and no disturbance variables.

Number of endogenous variables = 2
 Manifest: LAWYER ARCHTCT

Number of exogenous variables = 5

Manifest: EDUC VOCAB GEOMETRY
 Error: E_L E_A

The next part gives the initial values specified by the user. Because we did not specify any initial values, there are a lot of dots (SASese for missing values).

Manifest Variable Equations
 Initial Estimates

```

LAWYER =      .      *EDUC      +      .      *VOCAB      +      .      *GEOMETRY
              BLE              BLV              BLG
              + 1.0000 E_L

ARCHTCT =      .      *EDUC      +      .      *VOCAB      +      .      *GEOMETRY
              BAE              BAV              BAG
              + 1.0000 E_A
  
```

Variances of Exogenous Variables

Variable	Parameter	Estimate
EDUC	V_EDU	.
VOCAB	V_VOC	.
GEOMETRY	V_GEO	.
E_L	VE_L	.
E_A	VE_A	.

Covariances among Exogenous Variables

	Parameter	Estimate	
VOCAB	EDUC	COV_EV	.
GEOMETRY	EDUC	COV_EG	.
GEOMETRY	VOCAB	COV_VG	.
E_A	E_L	C_ELEA	.

Some descriptive information about the model along with descriptive statistics. The "informations" refer to the number of unique statistics in the observed covariance matrix, 15 in this case (5 variances and 10 covariances). The parameters are the number of unknowns. Because there are 15 parameters used to explain 15 statistics, there will be a perfect fit in this case. (That is, there are 15 equations in 15 unknowns.)

250 Observations	Model Terms	1
5 Variables	Model Matrices	4
15 Informations	Parameters	15

VARIABLE	Mean	Std Dev	
LAWYER	-0.01096000	1.039263660	Interests: lawyer
ARCHTCT	0.10920000	1.008410495	Interests: architect
EDUC	12.30400000	1.614313486	education in years
VOCAB	0.09016000	0.998324286	cognitive: vocabulary test
GEOMETRY	0.11252000	1.033761627	cognitive: geometry

Covariances

	LAWYER	ARCHTCT	EDUC
LAWYER	1.080068954	0.500160273	0.661497831
ARCHTCT	0.500160273	1.016891727	0.537914859
EDUC	0.661497831	0.537914859	2.606008032
VOCAB	0.440135897	0.284631855	0.835131888
GEOMETRY	0.297295200	0.213228530	0.690154538

	VOCAB	GEOMETRY	
LAWYER	0.4401358972	0.297295200	Interests: lawyer
ARCHTCT	0.2846318554	0.213228530	Interests: architect
EDUC	0.8351318876	0.690154538	education in years
VOCAB	0.9966513799	0.652774696	cognitive: vocabulary test
GEOMETRY	0.6527746956	1.068663102	cognitive: geometry

Determinant = 0.7651 (Ln = -0.268)

Ignore this next statement—a SAS bug.

Set covariances of exogenous manifest variables:

EDUC VOCAB GEOMETRY

Because we have not specified any start values, SAS automatically computes them for us.

Some initial estimates computed by two-stage LS method.

Vector of Initial Estimates

BLE	1	0.15509	Matrix Entry: <u>GAMMA</u> [1:1]
BLV	2	0.32513	Matrix Entry: <u>GAMMA</u> [1:2]
BLG	3	-0.02056	Matrix Entry: <u>GAMMA</u> [1:3]
BAE	4	0.15656	Matrix Entry: <u>GAMMA</u> [2:1]
BAV	5	0.14992	Matrix Entry: <u>GAMMA</u> [2:2]
BAG	6	0.00685	Matrix Entry: <u>GAMMA</u> [2:3]
V_EDU	7	2.60601	Matrix Entry: <u>PHI</u> [1:1]
COV_EV	8	0.83513	Matrix Entry: <u>PHI</u> [2:1]
V_VOC	9	0.99665	Matrix Entry: <u>PHI</u> [2:2]
COV_EG	10	0.69015	Matrix Entry: <u>PHI</u> [3:1]
COV_VG	11	0.65277	Matrix Entry: <u>PHI</u> [3:2]
V_GEO	12	1.06866	Matrix Entry: <u>PHI</u> [3:3]
VE_L	13	0.84049	Matrix Entry: <u>PHI</u> [4:4]
C_ELEA	14	0.32858	Matrix Entry: <u>PHI</u> [5:4]
VE_A	15	0.88855	Matrix Entry: <u>PHI</u> [5:5]

CALIS burps out some information here about the optimization algorithm that it is using. You can ignore this section.

Levenberg-Marquardt Optimization
 Scaling Update of More (1978)
 Number of Parameter Estimates 15
 Number of Functions (Observations) 15

Optimization Start: Active Constraints= 0 Criterion= 0.000
 Maximum Gradient Element= 0.000 Radius= 1.000

Iter rest nfun act optcrit difcrit maxgrad lambda rho

Optimization Results: Iterations= 0 Function Calls= 2 Jacobian Calls= 1
 Active Constraints= 0 Criterion= 0 Maximum Gradient Element= 5.02392E-16
 Lambda= 0 Rho= 0 Radius= 1

It is critical to read the NOTE given below. If the results *converge*, then CALIS has found a mathematical solution to the problem. If the message says that the algorithm failed to converge, then you should not trust the results

NOTE: ABSGCONV convergence criterion satisfied.

The covariance matrix predicted by the model.

Predicted Model Matrix

	LAWYER	ARCHTCT	EDUC
LAWYER	1.080068954	0.500160273	0.661497831
ARCHTCT	0.500160273	1.016891727	0.537914859
EDUC	0.661497831	0.537914859	2.606008032
VOCAB	0.440135897	0.284631855	0.835131888
GEOMETRY	0.297295200	0.213228530	0.690154538

	VOCAB	GEOMETRY	
LAWYER	0.4401358972	0.297295200	Interests: lawyer
ARCHTCT	0.2846318554	0.213228530	Interests: architect
EDUC	0.8351318876	0.690154538	education in years
VOCAB	0.9966513799	0.652774696	cognitive: vocabulary test
GEOMETRY	0.6527746956	1.068663102	cognitive: geometry

Determinant = 0.7651 (Ln = -0.268)

All the measures below are indices of how well the path model fits the data. In the present example, we actually have a perfect fit. Judging the fit of a model requires a combination of statistical acumen, considerable experience with structural equation modeling, and a lot of art. For now, interpret the row for χ^2 . If the model fits the data well, then the value of χ^2 relative to its degrees of freedom should be low and the p value should be high. When the χ^2 is large and the p value is small, then the path model should be rejected.

```

Fit criterion . . . . . 0.0000
Goodness of Fit Index (GFI) . . . . . 1.0000
GFI Adjusted for Degrees of Freedom (AGFI). . . . .
Root Mean Square Residual (RMR) . . . . . 0.0000
Parsimonious GFI (Mulaik, 1989) . . . . . 0.0000
Chi-square = 0.0000      df = 0      Prob>chi**2 = 0.0001
Null Model Chi-square:  df = 10      344.2237
RMSEA Estimate . . . . . 0.0000  90%C.I.[., .]
Probability of Close Fit . . . . .
ECVI Estimate . . . . . 0.1235  90%C.I.[., .]
Bentler's Comparative Fit Index . . . . . 1.0000
Normal Theory Reweighted LS Chi-square . . . . . 0.0000
Akaike's Information Criterion. . . . . 0.0000
Bozdogan's (1987) CAIC. . . . . 0.0000
Schwarz's Bayesian Criterion. . . . . 0.0000
McDonald's (1989) Centrality. . . . . 1.0000
Bentler & Bonett's (1980) Non-normed Index. . . . .
Bentler & Bonett's (1980) NFI . . . . . 1.0000
James, Mulaik, & Brett (1982) Parsimonious NFI. . . . . 0.0000
Z-Test of Wilson & Hilferty (1931). . . . .
Bollen (1986) Normed Index Rho1 . . . . .
Bollen (1988) Non-normed Index Delta2 . . . . . 1.0000
Hoelter's (1983) Critical N . . . . .

```

The residual matrix is the difference between the observed covariance matrix and the predicted covariance matrix. Because the path model used here gives a perfect fit to the data, all the residuals are 0.

```

Residual Matrix
      LAWYER  ARCHTCT  EDUC  VOCAB  GEOMETRY
LAWYER      0        0      0      0        0
ARCHTCT      0        0      0      0        0
EDUC         0        0      0      0        0
VOCAB        0        0      0      0        0
GEOMETRY     0        0      0      0        0

```

```

Average Absolute Residual = 0
Average Off-diagonal Absolute Residual = 0

```

Residuals in a covariance matrix give information about which elements are and which elements are not being predicted well by the model. However, what constitutes a "large" and "small" residual depends on the measurement scale of the variable. CALIS prints an estimate of standardized residuals that takes into account the measurement scale. Once again, all these elements are 0 because the model perfectly fits the data.

Asymptotically Standardized Residual Matrix

	LAWYER	ARCHTCT	EDUC	VOCAB	GEOMETRY
LAWYER	0	0	0	0	0
ARCHTCT	0	0	0	0	0
EDUC	0	0	0	0	0
VOCAB	0	0	0	0	0
GEOMETRY	0	0	0	0	0

Average Standardized Residual = 0

Average Off-diagonal Standardized Residual = 0

The unstandardized path coefficients, their standard errors, and the t value, along with the estimates of the variances and covariances among the exogenous variables. You should compare these path coefficients to the unstandardized regression coefficients from output from ~carey/p7291dir/pathreg1.sas where the same model was solved using multiple regression.

Manifest Variable Equations

LAWYER =	0.1551*EDUC	+ 0.3251*VOCAB	- 0.0206*GEOMETRY
Std Err	0.0424 BLE	0.0807 BLV	0.0732 BLG
t Value	3.6545	4.0305	-0.2810
	+ 1.0000 E_L		
ARCHTCT =	0.1566*EDUC	+ 0.1499*VOCAB	+ 0.0068*GEOMETRY
Std Err	0.0436 BAE	0.0829 BAV	0.0752 BAG
t Value	3.5879	1.8076	0.0910
	+ 1.0000 E_A		

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
EDUC	V_EDU	2.606008	0.233556	11.158
VOCAB	V_VOC	0.996651	0.089322	11.158
GEOMETRY	V_GEO	1.068663	0.095776	11.158
E_L	VE_L	0.840490	0.075327	11.158
E_A	VE_A	0.888546	0.079633	11.158

Covariances among Exogenous Variables

Parameter			Estimate	Standard Error	t Value
VOCAB	EDUC	COV_EV	0.835132	0.115030	7.260
GEOMETRY	EDUC	COV_EG	0.690155	0.114444	6.031
GEOMETRY	VOCAB	COV_VG	0.652775	0.077387	8.435
E_A	E_L	C_ELEA	0.328578	0.058590	5.608

The standardized solution. Once again, compare these path coefficients to the standardized regression weights on the output from ~carey/p7291dir/pathreg.sas.

Equations with Standardized Coefficients

$$\begin{aligned}
 \text{LAWYER} &= 0.2409 \cdot \text{EDUC} + 0.3123 \cdot \text{VOCAB} - 0.0205 \cdot \text{GEOMETRY} \\
 &\quad \text{BLE} \qquad \qquad \qquad \text{BLV} \qquad \qquad \qquad \text{BLG} \\
 &+ 0.8821 \cdot \text{E_L} \\
 \\
 \text{ARCHTCT} &= 0.2506 \cdot \text{EDUC} + 0.1484 \cdot \text{VOCAB} + 0.0070 \cdot \text{GEOMETRY} \\
 &\quad \text{BAE} \qquad \qquad \qquad \text{BAV} \qquad \qquad \qquad \text{BAG} \\
 &+ 0.9348 \cdot \text{E_A}
 \end{aligned}$$

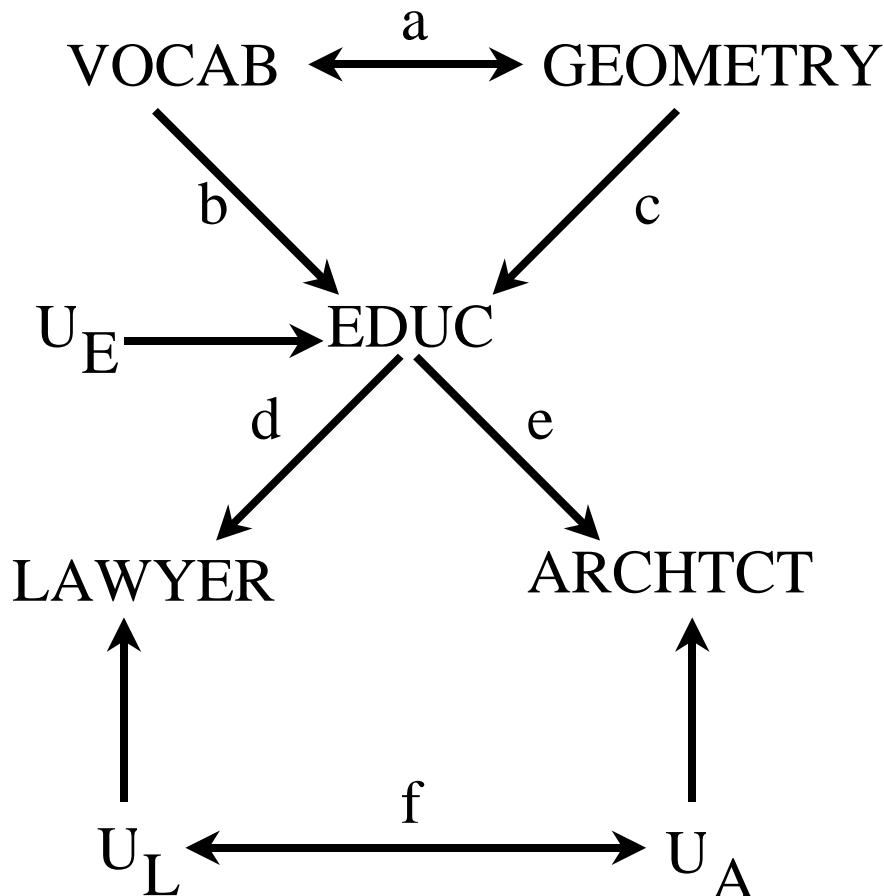
Squared Multiple Correlations

Variable	Error Variance	Total Variance	R-squared
1 LAWYER	0.840490	1.080069	0.221818
2 ARCHTCT	0.888546	1.016892	0.126214

Correlations among Exogenous Variables

Parameter			Estimate
VOCAB	EDUC	COV_EV	0.518198
GEOMETRY	EDUC	COV_EG	0.413560
GEOMETRY	VOCAB	COV_VG	0.632516
E_A	E_L	C_ELEA	0.380218

We now illustrate how CALIS can solve a model that cannot easily be solved using traditional multiple regression. The model is illustrated in the following path diagram.



This is known as a "mediational" model because EDUC is mediating the correlations between VOCAB and GEOMETRY on the one hand and LAWYER and ARCHTCT on the other hand. Substantively, this model implies that cognitive ability, as measured by VOCAB and GEOMETRY, allows one to have aspiration for and to actually achieve high levels of education. Both educational aspirations and achievements then predicts interests in these two professional occupations.

The SAS code for solving this problem is:

```

/* file ~carey/p7291dir/pathreg2a.sas. Same problem as
   ~carey/pathreg2.sas, but testing a mediational model */
PROC CALIS DATA=mvstats.interest COV CORR RESIDUAL;
LINEQS
    educ = b vocab + c geometry + e_e,
    lawyer = d educ + e_l,
    archtct = e educ + e_a;

STD vocab=v_v, geometry=v_g, e_e=ve_e, e_l=ve_l, e_a=ve_a;
COV vocab geometry=a, e_l e_a = f ;
RUN;

```

Selected SAS Output:

Note here that there are 11 parameters: the path coefficients *a* through *f* in the diagram, the variances of VOCAB and GEOMETRY, and the variances for the three error variables. The 11 parameters, along with their start values are listed below. Now, there are 15 equations in 11 unknowns so the model will not give a perfect fit to the data.

250 Observations	Model Terms	1
5 Variables	Model Matrices	4
15 Informations	Parameters	11

Vector of Initial Estimates

D	1	0.50675	Matrix Entry: <u>_BETA_[1:3]</u>
E	2	0.33411	Matrix Entry: <u>_BETA_[2:3]</u>
B	3	0.69167	Matrix Entry: <u>_GAMMA_[3:1]</u>
C	4	0.22331	Matrix Entry: <u>_GAMMA_[3:2]</u>
V_V	5	0.99665	Matrix Entry: <u>_PHI_[1:1]</u>
A	6	0.65277	Matrix Entry: <u>_PHI_[2:1]</u>
V_G	7	1.06866	Matrix Entry: <u>_PHI_[2:2]</u>
VE_E	8	1.87425	Matrix Entry: <u>_PHI_[3:3]</u>
VE_L	9	1.07885	Matrix Entry: <u>_PHI_[4:4]</u>
F	10	0.44778	Matrix Entry: <u>_PHI_[5:4]</u>
VE_A	11	0.94835	Matrix Entry: <u>_PHI_[5:5]</u>

Once again, it is crucial to check that the program has converged to a solution.

NOTE: ABSGCONV convergence criterion satisfied.

Here are the fit criteria. Notice that the χ^2 for 4 df is 20.64 and that it is significant (i.e., p value $< .05$). This means that the mediational model does not fit the data very well. Hence, the model should be rejected.

Fit criterion	0.0829	
Goodness of Fit Index (GFI)	0.9692	
GFI Adjusted for Degrees of Freedom (AGFI)	0.8843	
Root Mean Square Residual (RMR)	0.0751	
Parsimonious GFI (Mulaik, 1989)	0.3877	
Chi-square = 20.6417	df = 4	Prob>chi**2 = 0.0004
Null Model Chi-square:	df = 10	344.2237
RMSEA Estimate	0.1293	90%C.I.[0.0777, 0.1869]
Probability of Close Fit	0.0077	
ECVI Estimate	0.1734	90%C.I.[0.1305, 0.2473]
Bentler's Comparative Fit Index	0.9502	
Normal Theory Reweighted LS Chi-square	19.8139	
Akaike's Information Criterion.	12.6417	
Bozdogan's (1987) CAIC.	-5.4442	
Schwarz's Bayesian Criterion.	-1.4442	
McDonald's (1989) Centrality.	0.9673	
Bentler & Bonett's (1980) Non-normed Index.	0.8755	
Bentler & Bonett's (1980) NFI	0.9400	

James, Mulaik, & Brett (1982) Parsimonious NFI.	0.3760
Z-Test of Wilson & Hilferty (1931).	3.3246
Bollen (1986) Normed Index Rho1	0.8501
Bollen (1988) Non-normed Index Delta2	0.9511
Hoelter's (1983) Critical N	116

Notice both the residual matrix and the standardized residual matrix. There are certain statistics that the mediational model predicts perfectly, e.g., the variance of LAWYER, the covariance between LAWYER and ARCHTCT, etc. The model fails because it cannot predict the covariances between the two cognitive variables (VOCAB and GEOMETRY) and the two interest variables (LAWYER and ARCHTCT). Consequently, one would write this up by saying something to the effect that the model should be rejected because it underpredicts the correlations between the two cognitive variables and the two interest variables.

Residual Matrix

	LAWYER	ARCHTCT	EDUC
LAWYER	0.0000000000	0.0000000000	0.0000000000
ARCHTCT	0.0000000000	0.0000000000	0.0000000000
EDUC	0.0000000000	0.0000000000	0.0000000000
VOCAB	0.2281496233	0.1122494812	0.0000000000
GEOMETRY	0.1221093507	0.0707714169	0.0000000000

	VOCAB	GEOMETRY	
LAWYER	0.2281496233	0.1221093507	Interests: lawyer
ARCHTCT	0.1122494812	0.0707714169	Interests: architect
EDUC	0.0000000000	0.0000000000	education in years
VOCAB	0.0000000000	0.0000000000	cognitive: vocabulary test
GEOMETRY	0.0000000000	0.0000000000	cognitive: geometry

Average Absolute Residual = 0.03555
 Average Off-diagonal Absolute Residual = 0.05333

Rank Order of 4 Largest Residuals

VOCAB,LAWYER	GEOMETRY,LAWYER	VOCAB,ARCHTCT	GEOMETRY,ARCHTCT
0.228149623	0.122109351	0.112249481	0.070771417

Asymptotically Standardized Residual Matrix

	LAWYER	ARCHTCT	EDUC
LAWYER	0.0000000000	0.0000000000	0.0000000000
ARCHTCT	0.0000000000	0.0000000000	0.0000000000
EDUC	0.0000000000	0.0000000000	0.0000000000
VOCAB	4.414833979	2.179633761	0.0000000000
GEOMETRY	2.143504166	1.246630832	0.0000000000

	VOCAB	GEOMETRY	
LAWYER	4.414833979	2.143504166	Interests: lawyer
ARCHTCT	2.179633761	1.246630832	Interests: architect
EDUC	0.0000000000	0.0000000000	education in years
VOCAB	0.0000000000	0.0000000000	cognitive: vocabulary test
GEOMETRY	0.0000000000	0.0000000000	cognitive: geometry

Average Standardized Residual = 0.6656
 Average Off-diagonal Standardized Residual = 0.9985

Rank Order of 4 Largest Asymptotically Standardized Residuals
 VOCAB,LAWYER VOCAB,ARCHTCT GEOMETRY,LAWYER GEOMETRY,ARCHTCT
 4.414833979 2.179633761 2.143504166 1.246630832

The parameter estimates, first the unstandardized ones and then the standardized ones. If a model is rejected, one does not usually interpret the parameters.

Manifest Variable Equations

LAWYER = 0.2538*EDUC + 1.0000 E_L
 Std Err 0.0375 D
 t Value 6.7703

ARCHTCT = 0.2064*EDUC + 1.0000 E_A
 Std Err 0.0374 E
 t Value 5.5245

EDUC = 0.6917*VOCAB + 0.2233*GEOMETRY + 1.0000 E_E
 Std Err 0.1122 B 0.1084 C
 t Value 6.1646 2.0610

Variances of Exogenous Variables

Variable	Parameter	Estimate	Standard Error	t Value
VOCAB	V_V	0.996651	0.089322	11.158
GEOMETRY	V_G	1.068663	0.095776	11.158
E_E	VE_E	1.874248	0.167974	11.158
E_L	VE_L	0.912157	0.081749	11.158
E_A	VE_A	0.905859	0.081185	11.158

Covariances among Exogenous Variables

	Parameter		Estimate	Standard Error	t Value
GEOMETRY	VOCAB	A	0.652775	0.077387	8.435
E_A	E_L	F	0.363618	0.062044	5.861

Equations with Standardized Coefficients

LAWYER = 0.3943*EDUC + 0.9190 E_L
 D

ARCHTCT = 0.3304*EDUC + 0.9438 E_A
 E

EDUC = 0.4277*VOCAB + 0.1430*GEOMETRY + 0.8481 E_E
 B C

Squared Multiple Correlations

Variable		Error Variance	Total Variance	R-squared
1	LAWYER	0.912157	1.080069	0.155464
2	ARCHTCT	0.905859	1.016892	0.109188
3	EDUC	1.874248	2.606008	0.280797

Correlations among Exogenous Variables

Parameter			Estimate
GEOMETRY	VOCAB	A	0.632516
E_A	E_L	F	0.400019