Lab 21—Simple Path Analysis

Purpose

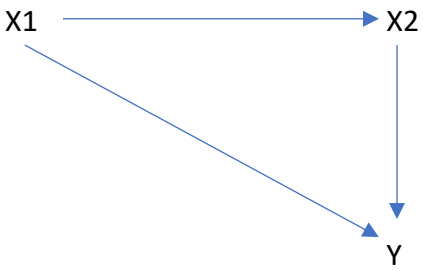
To build a simple path model.
To reconceptualize variables as exogenous and endogenous.
To distinguish direct and indirect relationships
To prepare graphic displays of path models

Main Points

Ordinary regression assumes that all the independent variables directly affect the dependent variable. In practice, however, variables often act indirectly through other variables.

As discussed in Lab 19a, one might theorize a basic developmental sequence among three variables such that X1→X2→Y.

A more sophisticated appreciation of the interrelationships among these variables might theorize that X1 affects Y thorough more than one path. For example, X1 may affect Y directly as well as indirectly via X2. Graphically this can be depicted as:



The approach of Path Analysis can be viewed as growing out of our discussion of statistical control. Path analysis extends the logic of statistical control in regression, focussing on interpretation (or mediation). The graphic above is essentially a slightly re-arranged depiction of partial mediation. In this approach it is common to speak of exogenous and endogenous variables rather than independent or dependent variables.

In the model presented here, X1 is considered exogenous since the source of its variation is outside of the model. Any variation in X1 is unexplained variance (UV). in this approach. The

model also depicts two endogenous variables, X2 and Y. Their variation is shown as due to the influence of other variables in the model. More complex models may have additional exogenous and endogenous variables. Path Analysis allows us to examine the direct and indirect effects of such models. Only models that flow in one direction (recursive) are considered here as those with feedback loops involve advanced statistical techniques.

Path analysis begins with a conceptual or graphic model setting out the causal relationships one thinks are at work in the world. Then the model is stated as a set of mathematical equations which are estimated using ordinary regression analysis. For the sake of completeness, such equations often explicitly include notations of unexplained variation (UV).

The three variable model depicted above can be summarized as:

$X_1 = UV$
$X_2 = b_1x_1 + UV.$
$Y= b_1x_1 + b_2x_2 + UV.$

These equations (other than the first) can be estimated using the regression procedure in SPSS. The number of regressions required depends upon the number of endogenous variables in the path model. To ensure completeness, it is perhaps helpful to proceed step by step "downstream" predicting each endogenous variable with all its upstream variables.

Example

This lab continues the example:
Partisanship→ideology→attitudes toward recreational marijuana. Partisanship is measured by Dem3, ideology by liberal5 and attitudes toward recreational marijuana by the index RawMJ3 (alpha =.777).

Syntax

```
*Regressions for Path Analysis—California 2016.
regression variables=RawMJ3 Democrat5 liberal5
  /statistics anova coeff r tol
  /descriptives = n
  /dependent = liberal5
  /method = enter Democrat5.

regression variables=RawMJ3 Democrat5 liberal5
  /statistics anova coeff r tol
  /descriptives = n
  /dependent = RawMJ3
  /method = enter Democrat5
```

/method = enter liberal5.

Output

Model 1:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.488[a]	.238	.238	.26727

a. Predictors: (Constant), Democrat5

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	.256	.017		14.899	<.001		
	Democrat5	.456	.026	.488	17.228	<.001	1.000	1.000

a. Dependent Variable: liberal5

Model 2:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.212[a]	.045	.044	1.12468
2	.352[b]	.124	.122	1.07746

a. Predictors: (Constant), Democrat5

b. Predictors: (Constant), Democrat5, liberal5

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Tolerance	VIF
1	(Constant)	1.108	.072		15.310	<.001		
	Democrat5	.743	.111	.212	6.668	<.001	1.000	1.000
2	(Constant)	.797	.077		10.348	<.001		
	Democrat5	.189	.122	.054	1.548	.122	.762	1.313
	liberal5	1.214	.131	.323	9.270	<.001	.762	1.313

a. Dependent Variable: RawMJ3

Displaying Results

It is generally helpful to display Path Analysis graphically. Preparing both standardized and standardized versions of the graphic offers flexibility in presenting results. This is because standardized coefficients are used to compare coefficients a single sample or dataset whereas unstandardized coefficients are more appropriate when comparing effects in different samples.

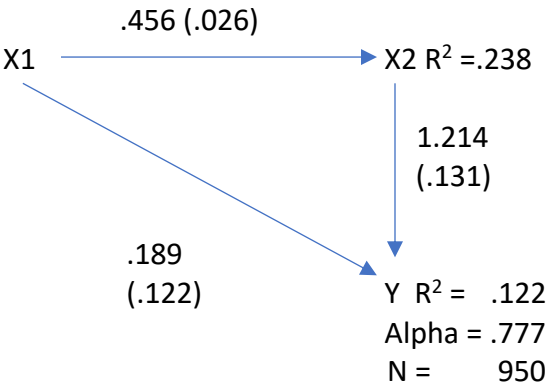
This is because there may be more variation in attitudes regarding recreational marijuana in one section of a country than in another. In such cases, the relative size of the standard regression coefficients may reflect the difference in variance rather than a difference in the relative strength or influence of the variables.

Coefficients produced by a path analysis that are not statistically significant are understood as not substantially different than zero. These coefficients can be eliminated from the diagram but need not be. Retaining an insignificant predictor in a diagram draws attention to the fact that it is not an effective predictor of a particular endogenous variable in the context of the multivariate model.

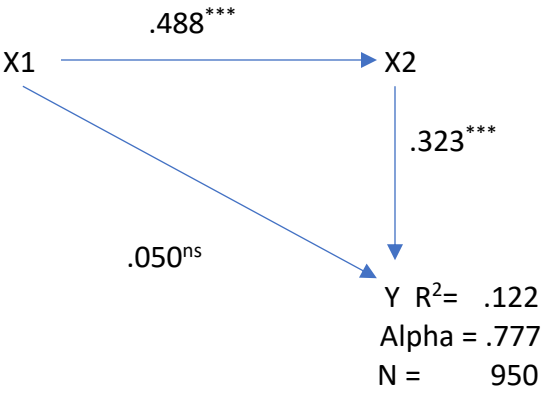
A twist in presenting the arrow diagram arises in that the proportion of variance explained can differ for each endogenous variable. Consequently, explained variance is indicated in terms of R^2 at more than one point in the model.

California 2016

Unstandardized Coefficients



Standardized Coefficients



X1= Dem5
X2= liberal5
Y= RawMJ3

Another advantage of path analysis is that it highlights both direct effects and indirect effects. Direct effects are already familiar and are illustrated by a direct arrow from one variable to another. They can be presented using either standardized or unstandardized coefficients. Indirect effects, however, are something new. An indirect effect is a path from one variable to another the involves more than one link. It is equal to the product of the direct paths of which is composed. Essentially one multiplies the coefficients together to estimate the indirect effect. Since units may differ substantially between one link in the direct path and another link, it is preferable to use standardized coefficients in computing them.

A variable's direct and indirect effects can be summed to show the total effects of one variable on another. A small table summarizing direct, indirect and total effects of one variable on another can be useful. In addition to standardized coefficients, this table can display the relative proportions of direct and indirect effects. Since the direct effect is not significant, the proportion of total effects column can be adjusted to reflect this.

Effects of Partisanship (Dem5) on Attitudes toward Marijuana Use (RawMJ3)

Type of Effect	Standardized Coefficient	Proportion of Total Effects
Direct	$.050^{ns}$	.245
Indirect	.154	.754
Total	.204	1.00

Source: derived from Figure 1

Interpretation of Results

This path analysis shows that the effect of partisanship on the dependent variable is indirect, in this case, completely mediated by ideology. This result is consistent with the finding of mediation (or interpretation) in Data Lab 19a. Here, the indirect effect, composed of paths from X1 to X2 and X2 to Y, is highlighted by the graphic display. This is evident by the regression coefficients displayed and the R^2 value displayed for X2 as well as Y. Moreover, this approach permits the calculation of direct, indirect and total effects. Not shown here, however, is the original significant relationship between X1 and Y, though a researcher could choose to include this in the graphic.

More substantively, this path analysis supports the view that Democratic partisanship leads respondents to develop a liberal ideological outlook which in turn leads to support for recreational marijuana use and legalization.