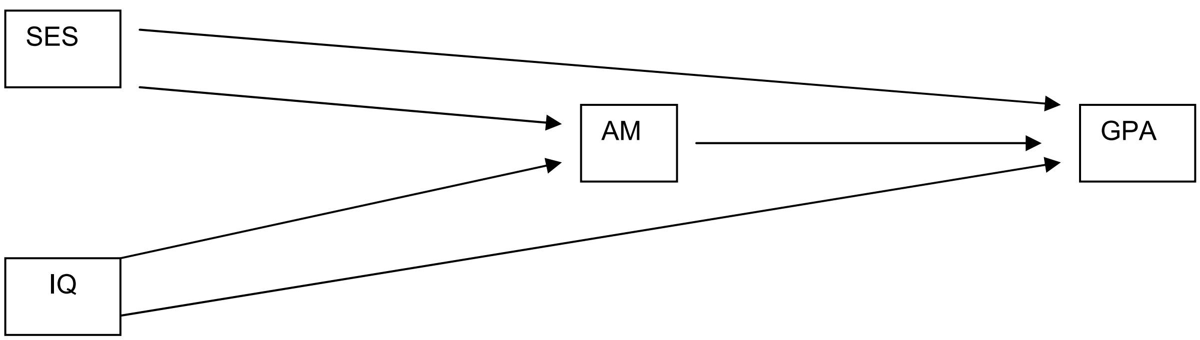
Example of Very Simple Path Analysis via Regression

Certainly the most three important sets of decisions leading to a path analysis are:

1. Which causal variables to include in the model
2. How to order the causal chain of those variables
3. Which paths are not "important" to the model – the only part that is statistically tested

Here's the hypothesized causal ordering for how SES, IQ & Achievement Motivation cause GPA. Usually a path analysis involves the analysis and comparison of two models – a "full model" with all of the possible paths included and a "reduced model" which has some of the paths deleted, because they are hypothesized to not contribute to the model.



The path coefficients for the full model (with all the arrows) are derived from a series of "layered" multiple regression analyses. For each multiple regression, the criterion is the variable in the box (all boxes after the leftmost layer) and the predictors are all the variables that have arrows leading to that box.

For the full model above, we will need two "layers" of multiple regressions:

1. with AM as the criterion and SES & IQ as the predictors
2. with GPA as the criterion and SES, IQ & AM as the predictors

The "First layer" multiple regression for the full model

Model Summary

Model	R	R Square
1	.412[a]	.169

a. Predictors: (Constant), IQ, SES

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.000	.172		.000	1.000
	SES	.616	.086	.398	7.177	.000
	IQ	8.810E-03	.012	.041	.734	.464

a. Dependent Variable: AM

The "Second Layer" multiple regression for the full model

Model Summary

Model	R	R Square
1	.705[a]	.496

a. Predictors: (Constant), AM, IQ, SES

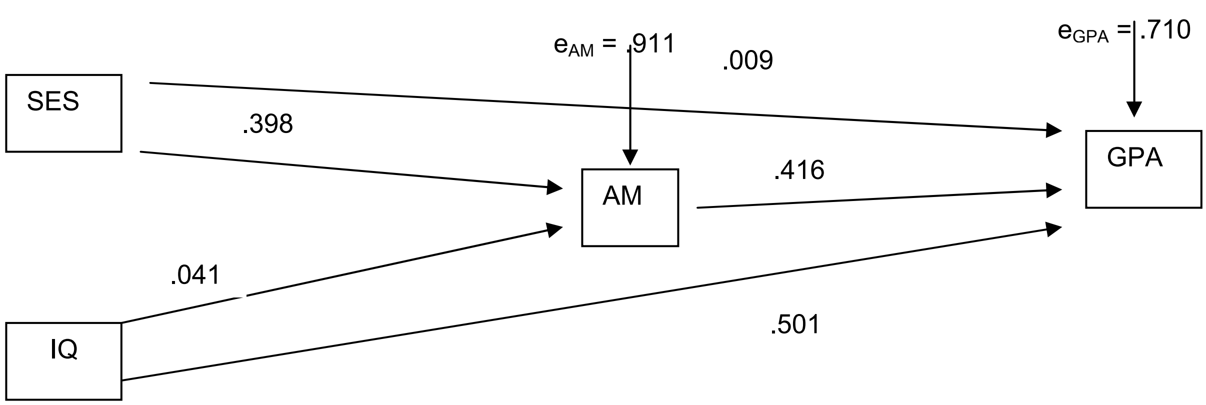
Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.000	.051		.000	1.000
	SES	5.470E-03	.028	.009	.196	.845
	IQ	4.172E-02	.004	.501	11.569	.000
	AM	.160	.017	.416	9.194	.000

a. Dependent Variable: GPA

Portraying the Full Path Model

- The path coefficients are the β weights from the multiple regression analyses.
- The "e" values (roughly error variance) are computed as $\sqrt{(1-R^2)}$ (e.g., $e_{AM} = \sqrt{(1-.169)} = .912$)

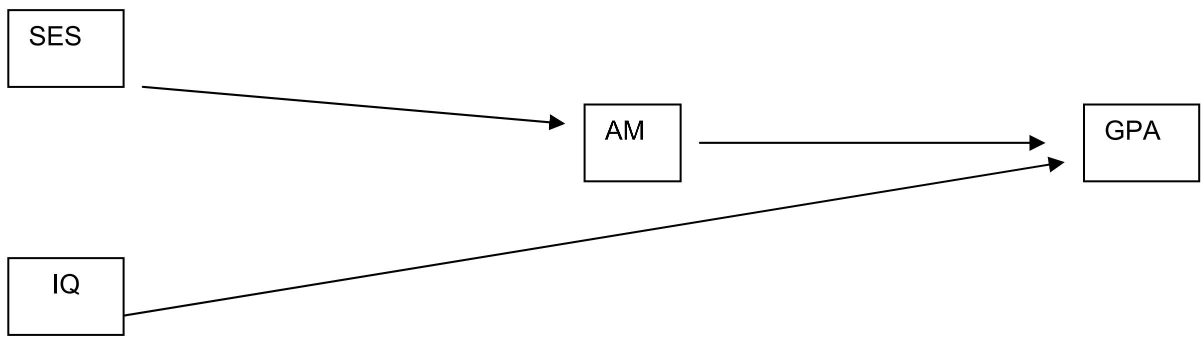


Examining this model we would note:

1. AM influences GPA
2. SES has no direct effect upon GPA, but has an indirect effect through AM
3. IQ has only a direct effect upon GPA

While some path analyses are "descriptive" in that they compute and describe this sort of "full model" others test hypotheses about which model paths do not portray causal links among the variables. Below is such a reduced model.

This model shows the research hypotheses that there is no direct effect of SES on GPA (that it's only effect is an indirect one channeled through AM) and IQ has only a direct effect (without any additional indirect effect channeled through AM).



Once again, two multiple regression models would be used to obtain the path coefficients.

The first layer doesn't require an actual multiple regression model, because there is only one predictor. So for AM as the criterion SES as the single predictor $R^2 = r^2 = .41^2 = .1681$, $\beta = r = .41$ and $e_{AM} = \sqrt{(1 - .1681)} = .911$

For the second layer we would find

Model Summary

Model	R	R Square
1	.705[a]	.496

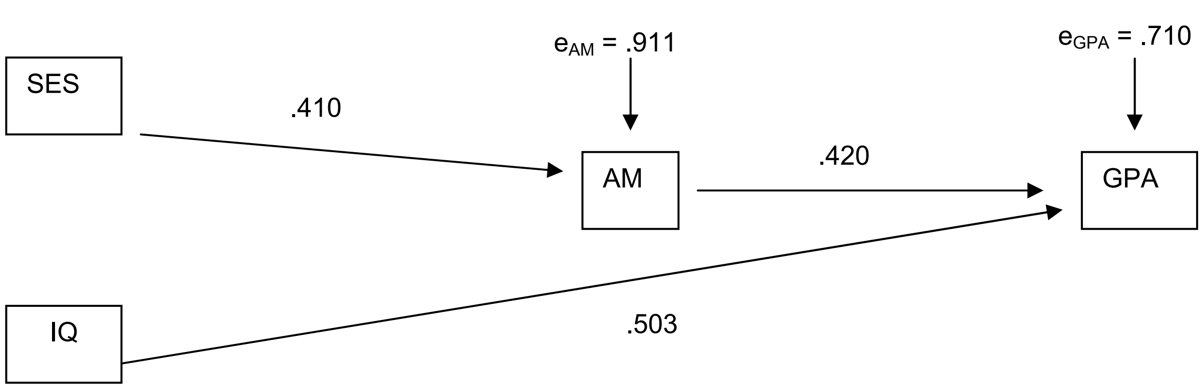
a. Predictors: (Constant), AM, IQ

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	.000	.051		.000	1.000
	IQ	4.191E-02	.003	.503	12.055	.000
	AM	.161	.016	.420	10.057	.000

a. Dependent Variable: GPA

Portraying the Reduced or Hypothesized Path Model



Testing the Reduced or Hypothesized Model

Testing the reduced model involves comparing how well it fits the data compared to how well the full model fits the data. This is much like the $R^2\Delta$ test for comparing nested models. As with those analyses, the test of the models actually tests the average contribution of the predictors (paths) being deleted from the model, so results from dropping several predictors can be uninformative or misleading.

Fit of the full model $\quad 1 - \pi(e^2) = 1 - .911^2 * .710^2 = .582$

Fit for the reduced model $\quad 1 - \pi(e^2) = 1 - .912^2 * .710^2 = .581$

The summary statistic showing the relative fit of the reduced model to the full model is

$$Q = \frac{1 - \text{fit of full model}}{1 - \text{fit of the reduced model}} = \frac{1 - .582}{1 - .581} = .9976$$

The significance test to compare the fit of the two models is (N = sample size d = number of dropped paths)

$$W = -(N - d) * \log_e Q = -(100 - 2) * \log_e .9976 = .235$$

W is distributed as X^2 with df = d. For this analysis $X^2(df=2, p = .05) = 5.991$. We would conclude that the reduced model fits the data as well as does the full model. That is, a causal model deleting the direct influence of SES and the indirect influence of IQ channeled through AM did not fit the data more poorly than did the model including these paths.