Regression And Mediation Analysis Using Mplus

Bengt Muthén & Mårten Schultzberg

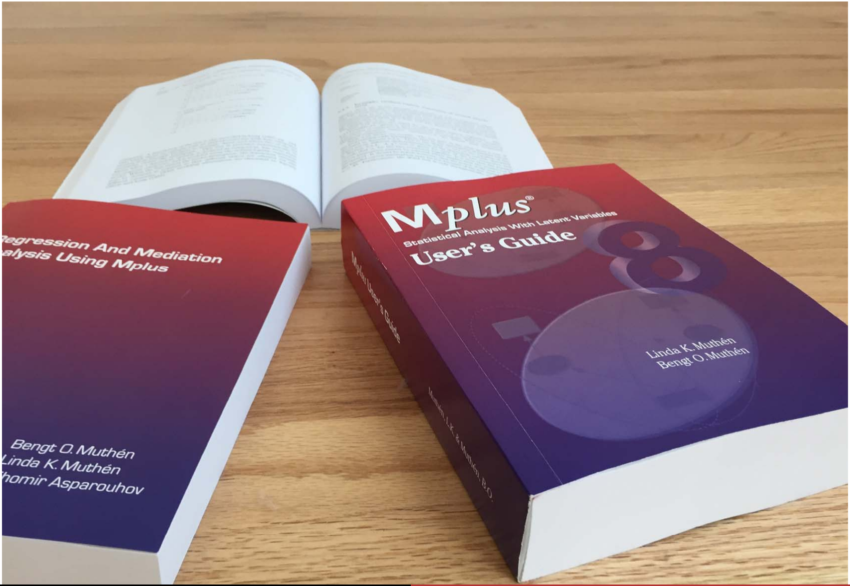
bmuthen@statmodel.com

Mplus
www.statmodel.com

Workshop at Johns Hopkins University
August 16, 2017

Expert assistance from Noah Hastings is acknowledged

The Mplus User's Guide has Gotten a Companion



- 1. Linear regression analysis
- 2. Mediation analysis
- 3. Special topics in mediation analysis
- 4. Causal inference for mediation
- 5. Categorical dependent variable
- 6. Count dependent variable
- 7. Censored dependent variable
- 8. Mediation with non-cont's variables
- 9. Bayesian analysis
- 10. Missing data

Table of Contents shown at www.statmodel.com. 500 pages. Published June 2016; third printing April 2017. Lots of inputs and outputs. Paperback. All inputs and outputs are posted. Most data sets are posted.

Overview of the Morning

- Block 1 (1 1/2 hours). Regression Analysis:
  - Linear regression with an interaction: A warm-up example
  - Categorical dependent variable: Not covered (prerequisite; book chapter 5)
  - Count dependent variable: Poisson, Poisson with a random intercept, zero-inflated Poisson, negative binomial, zero-inflated negbin, two-part (hurdle) model
  - Censored dependent variable: Tobit, censored-inflated, Heckman, and two-part analysis
- Block 2 (1 1/2 hours). Mediation Analysis (classic and modern):
  - Moderated mediation with continuous mediator and outcome
  - Monte Carlo simulation of moderated mediation
  - Sensitivity analysis
  - Modern mediation analysis using counterfactually-defined indirect and direct causal effects:
    - Binary mediator, binary outcome
    - Count outcome
    - Two-part outcome

Overview of the Afternoon

- Block 3 (1 1/2 hours). Mediation continued, Bayesian Analysis:
  - Prior, likelihood, posterior
  - Iterations, convergence, plots, model fit
  - Mediation examples: non-informative and informative priors
- Block 4 (1 1/2 hours). Bayesian Analysis continued, Missing Data Analysis:
  - Missing at random (MAR) maximum-likelihood estimation for regression and mediation
  - Missing on covariates: Benefits of using Bayes

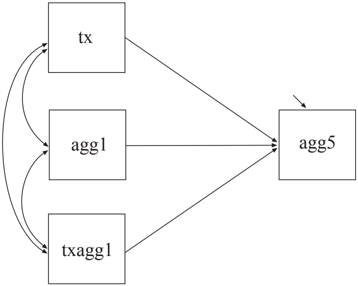
Ending at 4:45

15 minutes question and answer session at the end of each block

Outline

- **Introductory topics**
- Count dependent variable
- Censored dependent variable
- Classic mediation analysis
- Sensitivity analysis
- Modern mediation analysis
- Bayesian analysis
- Missing data analysis

Warmup Example: Linear Regression with an Interaction



Randomized field experiment in the Baltimore public schools where a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students was carried out (Kellam et al., 2008)

- tx is a binary intervention variable
- agg1 is pre-intervention Grade 1 aggressive behavior score and agg5 the score in Grade 5
- txagg1 is a treatment-baseline interaction (tx $\times$ agg1)

Example: Linear Regression with an Interaction

$$agg5_i = \beta_0 + \beta_1\, tx_i + \beta_2\, agg1_i + \beta_3\, txagg1_i + \varepsilon_i. \tag{1}$$

$$agg5_i = \beta_0 + \beta_1\, tx_i + \beta_2\, agg1_i + \beta_3\, tx_i\, agg1_i + \varepsilon_i \tag{2}$$

$$= \beta_0 + \beta_2\, agg1_i + (\beta_1 + \beta_3\, agg1_i)\, tx_i + \varepsilon_i. \tag{3}$$

- The expression $\beta_1 + \beta_3\, agg1$ is referred to as the moderator function
- Or, when evaluated at a specific agg1 value, the simple slope
- This means that agg1 moderates the β_1 effect of tx on agg5 by the term $\beta_3\, agg1$

Example: Input for Linear Regression with an Interaction

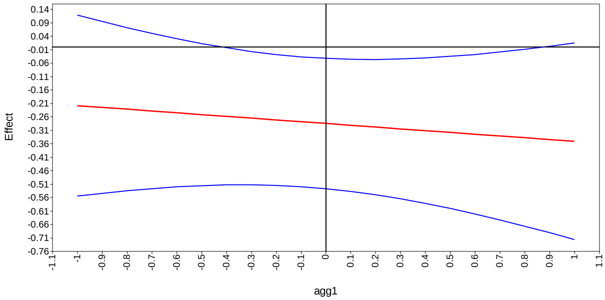
```
TITLE:              Linear regression with an interaction
DATA:               FILE = hopkins.dat;
VARIABLE:           NAMES = gender desgn11s sctaa15s sctaa11f;
                    USEVARIABLES = agg5 agg1 tx txagg1;
                    USEOBSERVATIONS = gender EQ 1 AND (desgn11s EQ 1 OR
                    desgn11s EQ 2 OR desgn11s EQ 3 OR desgn11s EQ 4);
DEFINE:             agg5 = sctaa15s/10;
                    agg1 = sctaa11f/10;
                    IF (desgn11s EQ 4) THEN tx=1;
                    IF (desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3) THEN
                    tx=0;
                    CENTER agg1(GRANDMEAN);
                    txagg1 = tx*agg1;
ANALYSIS:           ESTIMATOR = MLR;
MODEL:              agg5 ON
                    tx (b1)
                    agg1 (b2)
                    txagg1 (b3);
MODEL CONSTRAINT:
                    NEW(modlo mod0 modhi);
                    modlo = b1+b3*(-1.06);
                    mod0 = b1;
                    modhi = b1+b3*1.06;
OUTPUT:             SAMPSTAT PATTERNS STANDARDIZED RESIDUAL;
PLOT:               TYPE = PLOT3;
```

Example: Linear Regression with an Interaction

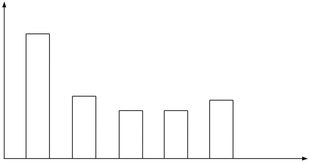
Table: Results for regression with a randomized intervention using treatment-baseline interaction ($n = 250$)

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
agg5 ON				
tx	-0.285	0.124	-2.307	0.021
agg1	0.500	0.076	6.543	0.000
txagg1	-0.066	0.130	-0.511	0.609
Intercepts				
agg5	2.483	0.077	32.238	0.000
Residual variances				
agg5	0.952	0.090	10.612	0.000
New/additional parameters				
modlo	-0.215	0.177	-1.211	0.226
mod0	-0.285	0.124	-2.307	0.021
modhi	-0.355	0.192	-1.849	0.064

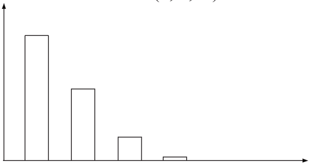
```
MODEL:      agg5 ON
            tx (b1)
            agg1 (b2)
            txagg1 (b3);
MODEL CONSTRAINT:
            LOOP(x,-1,1,0.1); ! moderator, lower limit, upper limit, increment
            PLOT(effect);
            effect = b1+b3*x;
```



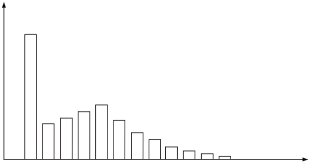
Categorical (e.g. strongly disagree,..., strongly agree)



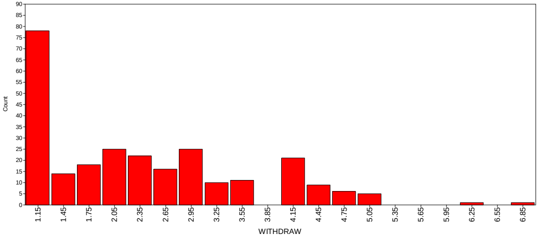
Count (0, 1, ...)



Censored (continuous)

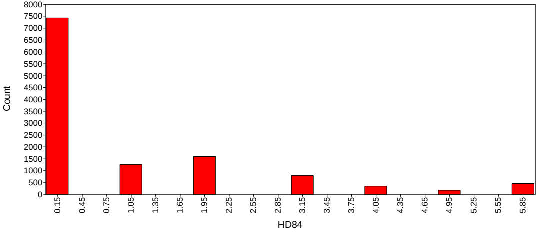


Distribution Example 1: Economic Stress (Hayes 2013)



- The outcome measures small-business owners' thoughts about withdrawing from their entrepreneurship due to economic stress
- Average of three 7-point items ranging from strongly disagree (1) to strongly agree (7)
- Participants were asked to rate if in the next year they would
  - "avoid entrepreneurial positions"
  - "feel anxious about entrepreneurial positions"
  - "feel less excited about entrepreneurial positions"

Distribution Example 2: Frequency of Heavy Drinking



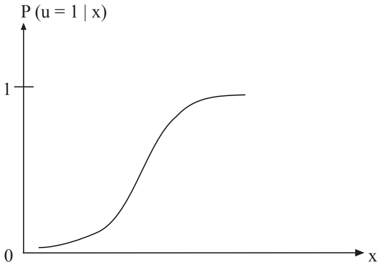
- "How often have you had 6 or more drinks on one occasion during the last 30 days?" (NLSY 1984)
  - Never (0)
  - Once (1)
  - 2 or 3 times (2)
  - 4 or 5 times (3)
  - 6 or 7 times (4)
  - 8 or 9 times (5)
  - 10 or more times (6)

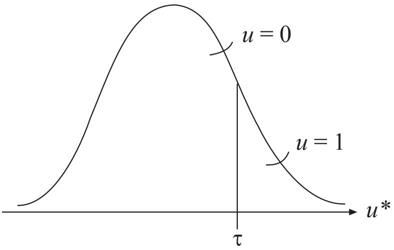
Categorical Variable Modeling

- Binary and ordinal variable: Logistic or probit
- Unordered categorical (nominal): Multinomial logistic

There are 3 common ways to describe binary variable (0/1) modeling with logistic regression:

- Probability: $\pi_i = P(u_i = 1|x_i) = \frac{e^{\beta_0+\beta_1\,x_i}}{1+e^{\beta_0+\beta_1\,x_i}} = \frac{1}{1+e^{-(\beta_0+\beta_1\,x_i)}}$
- Logit (log odds): $logit(\pi_i) = \beta_0 + \beta_1\,x_i$
- Latent response variable: $u_i^* = \beta_1\,x_i + \delta_i$





Outline

- Introductory topics
- **Count dependent variable**
- Censored dependent variable
- Classic mediation analysis
- Sensitivity analysis
- Modern mediation analysis
- Bayesian analysis
- Missing data analysis

Count Variable Modeling: Poisson Model

The Poisson distribution defines the probability of observing the count r for individual i on the count variable u_i as

$$P(u_i = r) = \frac{\mu_i^r \, e^{-\mu_i}}{r!}, \tag{4}$$

where $u_i = 0, 1, \ldots$ and μ is the mean also referred to as the rate at which the event occurs. The expression $r!$ is read as r factorial. For example, $3! = 1 \times 2 \times 3 = 6$. For $r = 0$, $r! = 1$.
Regression with a count dependent variable uses a linear regression for the log rate. The Poisson model specifies

$$log \; \mu_i = \beta_0 + \beta_1 \, x_i. \tag{5}$$

e^{β_1} is the change in the rate (mean) of u for a unit change in x

Poisson Model With A Random Intercept

The log rate model of (5) can be extended to a model with a residual ε that captures unobserved heterogeneity among individuals,

$$log\ \mu_i = \beta_0 + \beta_1\ x_i + \varepsilon_i, \tag{6}$$

where the residual is normally distributed, $\varepsilon \sim N(0, \sigma^2)$. The residual can be viewed as variation around the intercept β_0, interpreting the model as having a random intercept β_{0i},

$$\begin{aligned} log\ \mu_i &= \beta_{0i} + \beta_1\ x_i, \\ \beta_{0i} &= \beta_0 + \varepsilon_i. \end{aligned} \tag{7}$$

Expressed as (6), the residual can also be viewed as a continuous latent variable measured by the dependent variable $log\ \mu_i$ where the latent variable variance is an additional parameter to be estimated.

Zero-Inflated Poisson (ZIP)

- The Poisson model assumes that the variance of a Poisson variable is equal to its mean but count variables often have variances greater than the mean due to a preponderance of zeros
  - For example, in alcohol research, a common question format is "How many times in the last 30 days did you drink five or more drinks at one occasion?" A majority answers zero
- There are two reasons the answer zero is given
  - Some subjects may be non-drinkers and some subjects may be drinkers but have not engaged in heavy drinking during that period
  - In this way, a zero count is obtained through a mixture of two subpopulations or two latent classes, the zero class and the non-zero class
  - The term mixture is used because the class membership is not observed but is deduced from the data and the model

Let π denote the probability of being in the zero class (non-drinker in the alcohol example) so that $1-\pi$ is the probability of being in the class that follows a Poisson model where zero counts as well as positive counts can be observed.

Consider the mixture of the two classes for observing the count of zero

$$P(u_i = 0) = \pi_i + (1-\pi_i)\, e^{-\mu_i}, \tag{8}$$

where the first term represents the zero class and the second term represents the non-zero class where $e^{-\mu_i}$ is obtained from the Poisson distribution (4) when $u = 0$.

ZIP Model Continued

Allowing for a preponderance of zeros leads to the zero-inflated Poisson (ZIP) regression model which combines a logistic regression and a log rate equation,

$$logit(\pi_i) = \gamma_0 + \gamma_1 \, x_i, \tag{9}$$

$$log \, \mu_i = \beta_0 + \beta_1 \, x_i. \tag{10}$$

- The decision to not engage in the behavior is modeled differently than the extent of the behavior
- The binary dependent variable in (9) is unobserved and is referred to as a latent class variable in mixture modeling
- If the logit in (9) gets estimated at a large negative value, this implies a zero probability π of being in the zero class. In this case, the model is a regular Poisson model where the inflation part is not needed
- The mean for a zero-inflated Poisson regression model is $\mu_i(1 - \pi_i)$

Negative Binomial Model

The negative binomial model is expressed as

$$log\ \mu_i = \beta_0 + \beta_1\ x_i + \varepsilon_i, \tag{11}$$

where ε is a residual and e^{ε} has a gamma distribution. As in the random intercept Poisson model, the residual accounts for unobserved heterogeneity using a non-symmetric distribution.

- The negbin2 parameterization (see, e.g., Hilbe, 2011) has a mean of μ as does the Poisson model and a variance of $\mu(1 + \mu\ \alpha)$ where α is a dispersion parameter
- The Poisson model is obtained when $\alpha = 0$
- When $\alpha > 0$, the negative binomial model gives substantially higher probabilities for low counts and somewhat higher probabilities for high counts than the Poisson model

Other Count Models

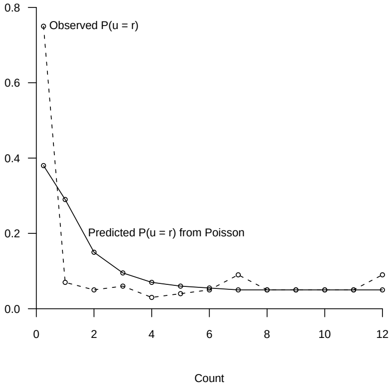
- Zero-inflated negbin
  - Binary and count part like for ZIP model
- Zero-truncated count
  - Zero probability for count = 0
- Hurdle (two-part)
  - Binary model for being at zero or not combined with zero-truncated count model
- Varying exposure
  - Length of observation time as offset (covariate with slope fixed = 1)

- Maximum-likelihood estimation
  - Not WLSMV
  - Not yet Bayes
- The Bayesian Information Criterion (BIC) is a useful statistic for comparing the different count models,

$$BIC = -2logL + r\ log\ n, \tag{12}$$

  where $logL$ is the maximized loglikelihood, r is the number of model parameters, and n is the sample size
  - A model with a lower BIC value is a better model in terms of balancing fit of the model to the data and model parsimony
  - The likelihood can be increased by increasing the number of parameters
  - BIC rewards models with high likelihood values and penalizes models with many parameters

- Dependent variable: number of marital affairs reported in the last year
- Covariates: having children, marital happiness, religiosity, and years married
- Sample size: $n = 601$



Input for Poisson Regression of Marital Affairs

```
TITLE:      Hilbe 2nd ed. page 248 example
DATA:       FILE = affairs1.dat;
VARIABLE:   NAMES = id male age yrsmarr kids relig educ occup ratemarr naff
            affair vryhap hapavg avgmarr unhap vryrel smerel slghtrel notrel;
            USEVAR= naffairs kids vryhap hapavg avgmarr vryrel smerel slgh
            notrel yrsmarr3 yrsmarr4 yrsmarr5 yrsmarr6;
            ! vryhap: very happily married
            ! hapavg: happily married
            ! avgmarr: avg marriage
            ! vryrel: very religious
            ! smerel: somewhat religious
            ! slghtrel: slightly religious
            ! notrel: not religious
            COUNT = naffairs; ! COUNT = naffairs (P);
DEFINE:     IF (yrsmarr==4) THEN yrsmarr3=1 ELSE yrsmarr3=0;
            IF (yrsmarr==7) THEN yrsmarr4=1 ELSE yrsmarr4=0;
            IF (yrsmarr==10) THEN yrsmarr5=1 ELSE yrsmarr5=0;
            if (yrsmarr==15) THEN yrsmarr6=1 ELSE yrsmarr6=0;
```

Input for Poisson Regression Continued

MODEL:	**naffairs ON kids-yrsmarr6;**
ANALYSIS:	ESTIMATOR = ML;
PLOT:	TYPE=PLOT3;

- The Poisson model with a random intercept is obtained by adding a latent variable f to the MODEL command,

```
MODEL:    naffairs ON kids-yrsmarr6;
          f BY naffairs;
```

  - The latent variable f represents the normally distributed residual ε in (6) with variance σ^2.

Input for Negbin and Inflation Models

- To get a negbin model, say: COUNT = naffairs(NB);

- To get a zero-inflated negbin model, say: COUNT = naffairs(NBI); and add a logistic model for the latent binary variable of being at zero or not:

```
MODEL:    naffairs ON kids-yrsmarr6;
          naffairs#1 ON kids-yrsmarr6;
```

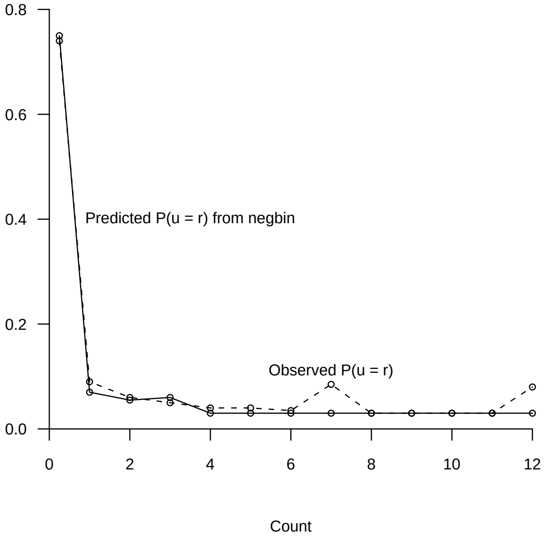
- Similarly, ZIP is obtained by: COUNT = naffairs(PI); again adding the logistic model for the binary variable

Output and Plots

COUNT PROPORTION OF ZERO, MINIMUM AND MAXIMUM VALUES			
NAFFAIRS	0.750	0	12

- Loglikelihood
- Parameter estimates, SEs, and z-tests
- Plots of observed proportions and estimated probabilities
    - Not conditioning on covariates: Estimated probability for x_i averaged over all individuals
    - Conditioned on covariates: Choose covariate values

Negbin Estimated Counts (Not Conditioning on Covariates)



Poisson and Negbin Estimates (Dispersion = 6.7, z=8.9)

	Results for Poisson regression of marital affairs			Results for negative binomial regression of marital affairs		
Parameter	Estimate	S.E.	Est./S.E.	Estimate	S.E.	Est./S.E.
naffairs ON						
kids	-0.223	0.106	-2.101	0.087	0.311	0.280
vryhap	-1.384	0.101	-13.707	-1.390	0.376	-3.688
hapavg	-1.024	0.086	-11.916	-0.980	0.365	-2.683
vgmarr	-0.886	0.105	-8.434	-0.971	0.429	-2.261
vryrel	-1.364	0.159	-8.579	-1.513	0.545	-2.778
smerel	-1.371	0.121	-11.300	-1.467	0.465	-3.157
slghtrel	-0.524	0.111	-4.407	-0.414	0.483	-0.857
notrel	-0.655	0.111	-5.894	-0.308	0.474	-0.649
yrsmarr3	0.758	0.161	4.701	0.668	0.398	1.681
yrsmarr4	1.105	0.170	6.502	1.335	0.446	2.993
yrsmarr5	1.480	0.165	8.979	1.189	0.448	2.653
yrsmarr6	1.480	0.156	9.515	1.427	0.387	3.686
Intercepts						
naffairs	1.102	0.165	6.684	0.816	0.626	1.303

Zero-Inflated Negbin: Inflation = P(being in the zero class)

	Count equation			Inflation equation		
Parameter	Estimate	S.E.	Est./S.E.	Estimate	S.E.	Est./S.E.
naffairs ON						
kids	-0.254	0.220	-1.154	-0.256	0.327	-0.783
vryhap	-0.445	0.226	-1.969	1.582	0.330	4.797
hapavg	-0.400	0.194	-2.063	1.091	0.309	3.537
avgmarr	-0.483	0.235	-2.060	0.774	0.357	2.166
vryrel	-0.569	0.343	-1.657	1.359	0.497	2.736
smerel	-0.461	0.273	-1.691	1.489	0.411	3.619
slghtrel	-0.096	0.259	-0.370	0.611	0.408	1.497
notrel	0.099	0.266	0.374	1.052	0.406	2.593
yrsmarr3	0.018	0.311	0.057	-0.762	0.408	-1.870
yrsmarr4	0.536	0.320	1.672	-0.652	0.451	-1.445
yrsmarr5	0.548	0.313	1.753	-0.986	0.457	-2.156
yrsmarr6	0.759	0.283	2.680	-0.754	0.407	-1.852
Intercepts						
naffairs	1.797	0.400	4.495	-0.269	0.518	-0.518
Dispersion						
naffairs	0.564	0.140	4.030			

Summary of Modeling Marital Affairs Data

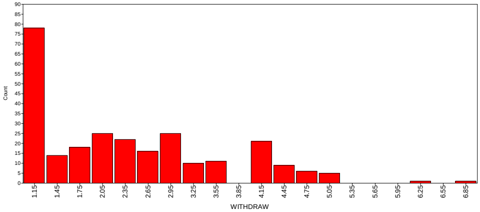
Model	Loglikelihood	#par.s	BIC
Poisson	-1,399.913	13	2883
Poisson with a random intercept	-735.942	14	1561
Negative binomial	-724.240	14	1538
Zero-inflated Poisson	-747.906	26	1652
Zero-inflated negative binomial	-689.718	27	1552
Two-part (hurdle)	-689.611	27	1552

Outline

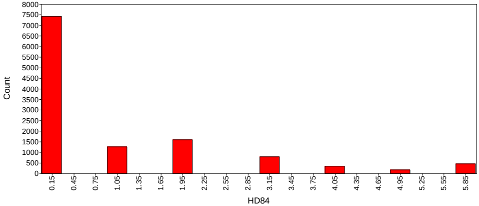
- Introductory topics
- Count dependent variable
- **Censored dependent variable**
- Classic mediation analysis
- Sensitivity analysis
- Modern mediation analysis
- Bayesian analysis
- Missing data analysis

Censored Variable Modeling

30% floor effect:

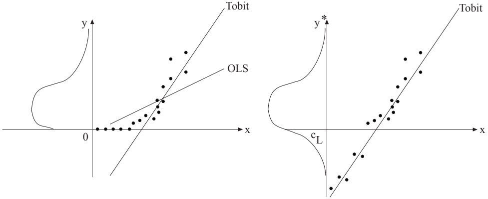


59% floor effect:



Regression Analysis Options in Mplus

- Censored-normal (Tobit)
- Censored-inflated
- Sample selection (Heckman)
- Two-part



$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{13}$$

$$y_i = \begin{cases} 0 & \text{if } y_i^* \leq 0 \\ y^* & \text{if } y_i^* > 0 \end{cases}$$

$$Binary\ (probit):\ P(y_i > 0 | x_i) = 1 - \Phi[\frac{0 - \beta_0 - \beta_1 x_i}{\sqrt{V(\varepsilon)}}] = \Phi[\frac{\beta_0 + \beta_1 x_i}{\sqrt{V(\varepsilon)}}], \tag{14}$$

$$Continuous, positive:\ E(y_i | y_i > 0, x_i) = \beta_0 + \beta_1\, x_i + \sqrt{V(\varepsilon)}\, \frac{\phi(z_i)}{\Phi(z_i)}, \tag{15}$$

- Latent class 0: subjects for whom only $y = 0$ is observed
- Latent class 1: subjects following a censored-normal (tobit) model

Assume a logistic regression that describes the probability of being in class 0,

$$logit(\pi_i) = \gamma_0 + \gamma_1\, x_i. \tag{16}$$

For subjects in class 1 the usual censored-normal model of (17) is assumed with

$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i. \tag{17}$$

Two ways $y = 0$ is observed (mixture at zero).

Consider the linear regression for the continuous latent response variable y^*,

$$y_i^* = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{18}$$

where the latent response variable y_i^* is observed as $y_i = y_i^*$ when a binary variable $u_i = 1$ and remains latent, that is, missing if $u_i = 0$. A probit regression is specified for u,

$$u_i^* = \gamma_1\, x_i + \delta_i, \tag{19}$$

where the categories of the binary observed variable u_i are determined by u^* falling below or exceeding a threshold parameter τ,

$$u_i = \begin{cases} 0 & \text{if } u_i^* \leq \tau \\ 1 & \text{if } u_i^* > \tau. \end{cases}$$

A key feature is that the residuals ε and δ are assumed to be correlated and have a bivariate normal distribution with the usual probit standardization $V(\delta) = 1$.

Two-Part Regression

With censoring from below at zero and using probit regression with the event of $u = 1$ referring to a positive outcome, the two-part model is expressed as

$$probit(\pi_i) = \gamma_0 + \gamma_1\, x_i, \tag{20}$$

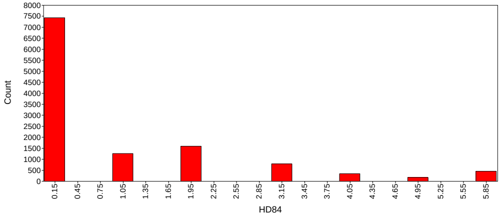
$$log\ y_{i|u_i=1} = \beta_0 + \beta_1\, x_i + \varepsilon_i, \tag{21}$$

where $\pi_i = P(u_i = 1|x_i)$ and $\varepsilon_i \sim N(0, V(\varepsilon))$. Logistic regression can be used as an alternative to the probit regression in (20). Maximum-likelihood estimation of the two-part model gives the same estimates as if the binary and the continuous parts were estimated separately using maximum-likelihood. Expressing (20) in terms of a latent response variable regression with a normal residual, the two residuals can be correlated but the correlation does not enter into the likelihood and is not estimated.

Comparison of Censored-Inflated, Heckman, and Two-Part

- Like the censored-inflated and Heckman models, the two-part model has different regression equations for the two parts
- Unlike the censored-inflated model, the two-part model does not have a mixture at zero, nor does Heckman
- Unlike the Heckman model, the two-part model does not estimate a residual correlation between the two parts
- Duan et al. (1983) pointed to two advantages of the two-part model over Heckman:
  - Applied to medical care expenses, it is preferable to the Heckman model because the censoring point of zero expense does not represent missing data but rather a real, observed value
  - A bivariate normality assumption for the residuals is not needed

Example: Comparing Methods on Heavy Drinking Data



NLSY Data on Heavy Drinking ($n = 1,152$)

- Dependent variable: frequency of heavy drinking measured by the question:
  - "How often have you had 6 or more drinks on one occasion during the last 30 days?"
  - Never (0); once (1); 2 or 3 times (2); 4 or 5 times (3); 6 or 7 times (4); 8 or 9 times (5); and 10 or more times (6)
- Covariates: gender, ethnicity, early onset of regular drinking (es), family history of problem drinking, and high school dropout.

Input for Censored-Normal (Tobit) and Censored-Inflated

```
           USEVARIABLES = hd84 male black hisp es fh123 hsdrp;
           CENSORED = hd84 (B);
ANALYSIS:  ESTIMATOR = MLR;
MODEL:     hd84 ON male black hisp es fh123 hsdrp;
```

```
           USEVARIABLES = hd84 male black hisp es fh123 hsdrp;
           CENSORED = hd84 (BI);
ANALYSIS:  ESTIMATOR = MLR;
MODEL:     hd84 ON male black hisp es fh123 hsdrp;
           hd84#1 ON male black hisp es fh123 hsdrp;
```

DATA TWOPART

The DATA TWOPART command is used to create a binary and a continuous variable from a continuous variable with a floor effect. A cutpoint of zero is used as the default. Following are the rules used to create the two variables:

1. If the value of the original variable is missing, both the new binary and the new continuous variable values are missing
2. If the value of the original variable is greater than the cutpoint value, the new binary variable value is one and the new continuous variable value is the log of the original variable as the default
3. If the value of the original variable is less than or equal to the cutpoint value, the new binary variable value is zero and the new continuous variable value is missing

Input for Heckman and Two-Part

	USEVARIABLES = male black hisp es fh123 hsdrp **u positive;**
	CATEGORICAL = u;
DATA TWOPART:	
	NAMES = hd84;
	BINARY = u;
	CONTINUOUS = positive;
ANALYSIS:	ESTIMATOR = MLR;
	LINK = PROBIT;
	MCONVERGENCE = 0.00001;
	INTEGRATION = 30; ! See Lesaffre & Spiessens (2001) Appl Stat
MODEL:	**positive u ON male black hisp es fh123 hsdrp;**
	f BY u positive; f@1;

	USEVARIABLES = male black hisp es fh123 hsdrp **u positive;**
	CATEGORICAL = u;
DATA TWOPART:	
	NAMES = hd84;
	BINARY = u;
	CONTINUOUS = positive;
ANALYSIS:	ESTIMATOR = MLR;
	LINK = PROBIT;
MODEL:	**positive u ON male black hisp es fh123 hsdrp;**
OUTPUT:	TECH1 TECH8;

Loglikelihood and BIC for Four Models for Frequency of Heavy Drinking

The Heckman and two-part models use $log(y)$ so logL and BIC values cannot be compared to those of tobit and censored-inflated:

Model	log L	# parameters	BIC
Censored-normal (tobit)	-1530.512	8	3117
Censored-inflated	-1499.409	15	3105
Sample selection (Heckman)	-1088.182	16	2289
Two-part	-1088.400	15	2283

Results for the Censored-Normal (Tobit) Regression Model

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
hd84 ON				
male	2.106	0.210	10.038	0.000
black	-2.157	0.258	-8.359	0.000
hisp	-1.059	0.298	-3.555	0.000
es	0.716	0.286	2.503	0.012
fh123	0.615	0.317	1.938	0.053
hsdrp	0.240	0.265	0.908	0.364
Intercepts				
hd84	-1.258	0.211	-5.961	0.000
Residual variances				
hd84	8.678	0.559	15.525	0.000

Results for the Censored-Inflated Regression Model

Parameter	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
hd84 ON				
male	0.957	0.236	4.057	0.000
black	-1.150	0.282	-4.073	0.000
hisp	-0.405	0.320	-1.264	0.206
es	0.585	0.276	2.120	0.034
fh123	-0.031	0.329	-0.095	0.924
hsdrp	0.390	0.263	1.487	0.137
hd84#1 ON				
male	-1.025	0.166	-6.157	0.000
black	0.962	0.208	4.621	0.000
hisp	0.570	0.215	2.651	0.008
es	-0.204	0.198	-1.032	0.302
fh123	-0.512	0.273	-1.876	0.061
hsdrp	0.040	0.188	0.213	0.831
Intercepts				
hd84#1	0.412	0.145	2.848	0.004
hd84	1.567	0.189	8.290	0.000

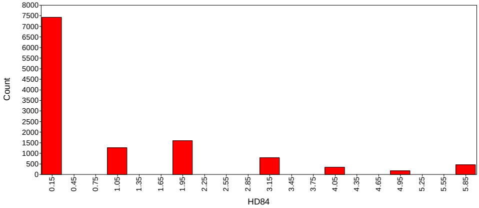
Comparisons of Results

- Heckman versus Two-part:
  - Very similar logL/BIC and results (the Heckman probit coefficients need to be divided by $\sqrt{2}$ due to adding the factor)
  - The Heckman residual correlation is significant
- Censored-inflated versus Two-part:
  - Similar results (reverse signs for the binary part)
  - LogL and BIC not comparable but limited model fit comparison can be made using MODEL CONSTRAINT:

Table: Estimated probability of zero heavy drinking and mean of heavy drinking for a subset of males who have zero values on the covariates black, hisp, es, fh123, and hsdrp

	Probability	Mean
Sample values	0.441	1.538
Censored-inflated estimates	0.402	1.547
Two-part estimates	0.403	1.671

Heckman and Two-Part Treating the Positive Part as Ordinal



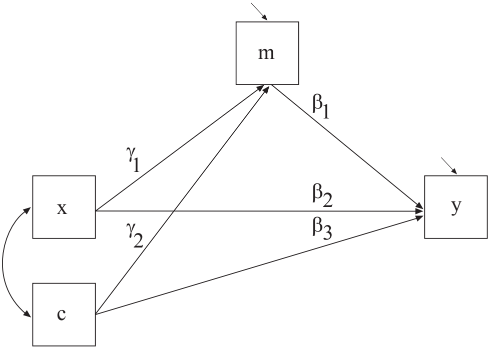
- Assignment: As an alternative, an ordinal approach may be good for these data given
  1. the limited number of response categories
  2. the slight ceiling effect for category 6, 10 or more times so that the assumption of a log normal distribution can be questioned:

  - Declare the positive part as categorical using the CATEGORICAL option of the VARIABLE command
  - Use TRANSFORM = NONE in the DATA TWOPART command to avoid the log transformation

Outline

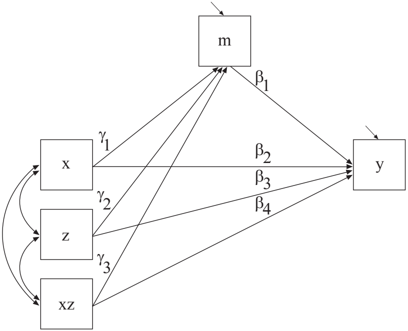
- Introductory topics
- Count dependent variable
- Censored dependent variable
- **Classic mediation analysis**
- Sensitivity analysis
- Modern mediation analysis
- Bayesian analysis
- Missing data analysis

Figure: A basic mediation model with an exposure variable x, a control variable c, a mediator m, and an outcome y



Moderated Mediation Analysis: Case 1 (xz)

Figure: Case 1 moderated mediation of y on x, m on x, both moderated by z

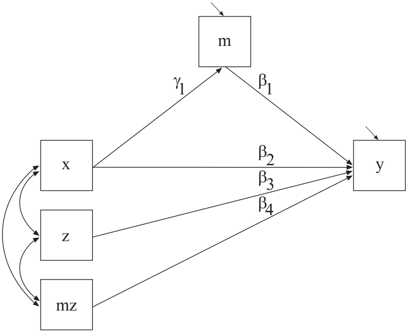


$$Indirect: \beta_1\,(\gamma_1 + \gamma_3\,z)(x_1 - x_0), \tag{22}$$

$$Direct: (\beta_2 + \beta_4\,z)(x_1 - x_0). \tag{23}$$

$x_1 - x_0$ often represents a one-unit change or a change from 0 to 1

Figure: Case 2 moderated mediation of *y* on *m* moderated by *z*

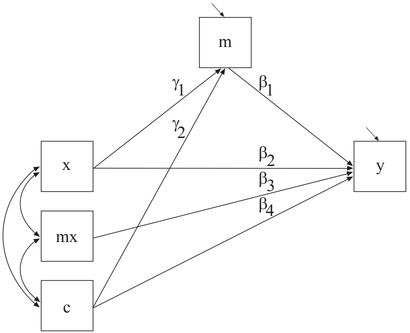


$$Indirect: (\beta_1 + \beta_4\, z)\gamma_1(x_1 - x_0), \tag{24}$$

$$Direct: \beta_2(x_1 - x_0). \tag{25}$$

Moderated Mediation Analysis: Case 3 (*mx*)

Figure: Case 3 moderated mediation of *y* on *m* moderated by *x*



$$Indirect: (\beta_1 + \beta_3\, x_1)\gamma_1(x_1 - x_0), \tag{26}$$

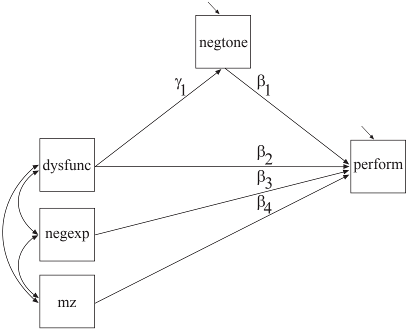
$$Direct: (\beta_2 + \beta_3(\gamma_0 + \gamma_1\, x_0 + \gamma_2\, c))(x_1 - x_0). \tag{27}$$

Mplus Options for Moderated Mediation (Single M; Also for Counterfactually-Defined Effects)

- No Moderation:
  - Y IND M X.
  - All 3 can be latent.
- Moderation with Z:
  - Involving X (4 arguments after MOD; **Case 1**):
    - Y MOD M Z(low, high, increment) XZ Z;
    - M and Y can be latent.
  - Involving M (4 arguments after MOD; **Case 2**):
    - Y MOD M Z(low, high, increment) MZ X;
    - X and Y can be latent.
  - Involving X and M (5 arguments after MOD):
    - Y MOD M Z(low, high, increment) MZ XZ X,
    - Only Y can be latent.
- Moderation with M*X (3 arguments after MOD; **Case 3**):
  - y MOD M MX X;
  - Y can be latent.

Example: Case 2 Moderated Mediation for Work Team Performance (Hayes, 2013; $n = 60$)

Figure: Case 2 (*mz*) moderated mediation for work team behavior. The exposure variable is dysfunc (continuous). The interaction variable *mz* is the product of the mediator variable negtone and the moderator variable negexp

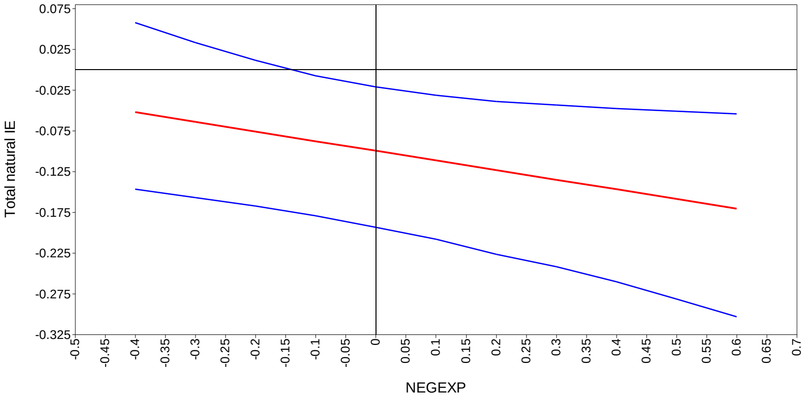


Input for Case 2 Moderated Mediation for Work Teams

```
TITLE:            Hayes (2013) TEAMS Case 2 moderation of M ->Y
DATA:             FILE = teams.txt;
VARIABLE:         NAMES = dysfunc negtone negexp perform;
                  USEVARIABLES = dysfunc negtone negexp perform mz;
DEFINE:           mz = negtone*negexp;
ANALYSIS:         ESTIMATOR = ML;
                  BOOTSTRAP = 10000;
MODEL:            perform ON negtone dysfunc negexp mz;
                  negtone ON dysfunc;
MODEL INDIRECT:
                  perform MOD negtone negexp(-.4,.6,.1)
                  mz dysfunc(0.4038 0.035);
OUTPUT:           SAMPSTAT CINTERVAL(BOOTSTRAP);
PLOT:             TYPE = PLOT3;
```

- The moderator variable negexp has 20^{th} and 80^{th} percentiles -0.4 and 0.6, respectively
- The exposure variable dysfunc has mean 0.4038 and standard deviation 0.369 so that $x_1 - x_0 = 0.4038 - 0.035 = 0.369$. In other words, 0.035 is one standard deviation below the mean

Figure: Indirect effect and bootstrap confidence interval for case 2 (*mz*) moderated mediation for work team behavior. The moderator variable is negexp and the indirect effect is labeled Total natural IE

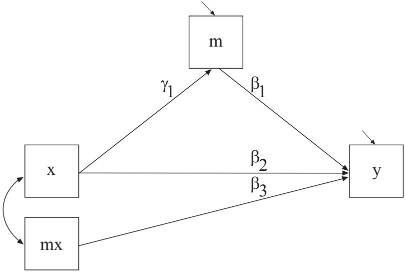


Ignore Chi-Square Test of Model Fit When Interaction Involves the Mediator

An alternative specification used in Preacher et al. (2007) avoids the two degrees of freedom that arise because of the two left-out arrows in the model. This saturates the model by allowing covariances between the moderator variable and the mediator residual and between the moderator-exposure interaction variable and the mediator residual. To accomplish this, the MODEL specification adds a line using WITH:

```
MODEL:
        perform ON negtone dysfunc negexp mz;
        negtone ON dysfunc;
        negexp mz WITH negtone dysfunc;
```

No change in estimates or SEs if covariances are not included.



The effects of x on y are

$$Indirect : (\beta_1 + \beta_3\, x_1)\gamma_1(x_1 - x_0), \tag{28}$$

$$Direct : (\beta_2 + \beta_3(\gamma_0 + \gamma_1\, x_0))(x_1 - x_0). \tag{29}$$

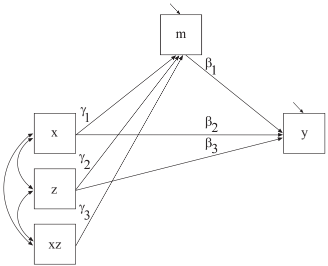
Non-Significant Exposure-Mediator Interaction

Quoting VanderWeele (2015, p. 46):

> *"An investigator might be tempted to only include such exposure-mediator interactions in the model if the interaction is statistically significant. - - This approach is problematic. It is problematic because power to detect interaction tends to be very low unless the sample size is very large. - - such exposure-mediator interaction may be important in capturing the dynamics of mediation... - - A better approach - - is perhaps to include them by default and only exclude them if they do not seem to change the estimates of the direct and indirect effects very much."*

Input for Case 3 Moderated Mediation of Simulated Data

TITLE:	x moderation of y regressed on m
DATA:	FILE = xmVx4s1n200rep6.dat;
VARIABLE:	NAMES = y m x;
	USEVARIABLES = y m x mx;
DEFINE:	**mx = m*x;**
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
MODEL:	**y ON m x mx;**
	m ON x;
MODEL INDIRECT:	
	y MOD m mx x(7 5);
OUTPUT:	SAMPSTAT CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;



The model used for data generation is

$$y_i = \beta_0 + \beta_1\, m_i + \beta_2\, x_i + \beta_3\, z_i + \varepsilon_{yi}, \tag{30}$$

$$m_i = \gamma_0 + \gamma_{1i}\, x_i + \gamma_2\, z_i + \varepsilon_{mi}, \tag{31}$$

$$\gamma_{1i} = \gamma_1 + \gamma_3\, z_i, \tag{32}$$

where γ_{1i} is a random slope. Inserting (32) in (31) shows that the random slope formulation is equivalent to adding an interaction term xz as a covariate in the regression of m.

```
TITLE:              Simulating Z moderation of X to M using a random slope, saving the
                    data for external Monte Carlo analysis
MONTECARLO:
                    NAMES = y m x z;
                    NOBS = 400;
                    NREPS = 500;
                    REPSAVE = ALL;
                    SAVE = xzrep*.dat;
                    CUTPOINTS = x(0);
MODEL POPULATION:
                    x-z@1; [x-z@0];
                    x WITH z@0.5;
                    y ON m*.5 x*.2 z*.1; y*.5; [y*0];
                    gamma1 | m ON x;
                    [gamma1*.3];
                    gamma1 ON z*.2;
                    gamma1@0;
                    m ON z*.3; m*1; [m*0];
ANALYSIS:           TYPE = RANDOM;
MODEL:              y ON m*.5 (b)
                    x*.2 z*.1;
                    y*.5; [y*0];
                    gamma1 | m ON x;
                    [gamma1*.3] (gamma1);
                    gamma1 ON z*.2 (gamma3);
                    gamma1@0;
                    m ON z*.3; m*1; [m*0];
MODEL CONSTRAINT:
                    NEW(indavg*.15 indlow*.05 indhigh*.25);
                    indavg = b*gamma1;
                    indlow = b*(gamma1-gamma3);
                    indhigh = b*(gamma1+gamma3);
```

Results for Monte Carlo Simulation of z Moderation of m Regressed on x using $n = 400$ and 500 Replications

	Population	Average	Std. Dev.	S.E. Average	M.S.E.	95% Cover	% Sig Coeff
gamma1 ON							
z	0.200	0.2010	0.0775	0.0771	0.0060	0.950	0.744
y ON							
m	0.500	0.5007	0.0524	0.0494	0.0027	0.922	1.000
x	0.200	0.2056	0.0783	0.0784	0.0061	0.938	0.754
z	0.100	0.0963	0.0470	0.0433	0.0022	0.926	0.604
m ON							
z	0.300	0.2999	0.0531	0.0545	0.0028	0.964	1.000
Intercepts							
y	0.000	-0.0017	0.0527	0.0522	0.0028	0.934	0.066
m	0.000	-0.0008	0.0543	0.0545	0.0029	0.946	0.054
gamma1	0.300	0.3010	0.0776	0.0770	0.0060	0.962	0.978
Residual Variances							
y	0.500	0.4938	0.0341	0.0347	0.0012	0.928	1.000
m	0.500	0.4940	0.0331	0.0346	0.0011	0.950	1.000
gamma1	0.000	0.0000	0.0000	0.0000	0.0000	1.000	0.000
New/Additional Parameters							
indavg	0.150	0.1505	0.0417	0.0416	0.0017	0.956	0.974
indlow	0.050	0.0497	0.0546	0.0548	0.0030	0.958	0.138
indhigh	0.250	0.2514	0.0628	0.0603	0.0039	0.928	0.988

- Introductory topics
- Count dependent variable
- Censored dependent variable
- Classic mediation analysis
- **Sensitivity analysis**
- Modern mediation analysis
- Bayesian analysis
- Missing data analysis

Figure: Mediator-outcome confounding 1

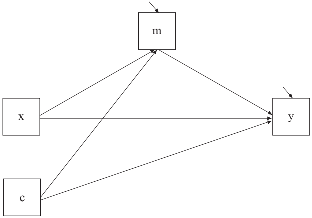
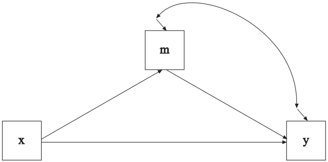


Figure: Mediator-outcome confounding 2

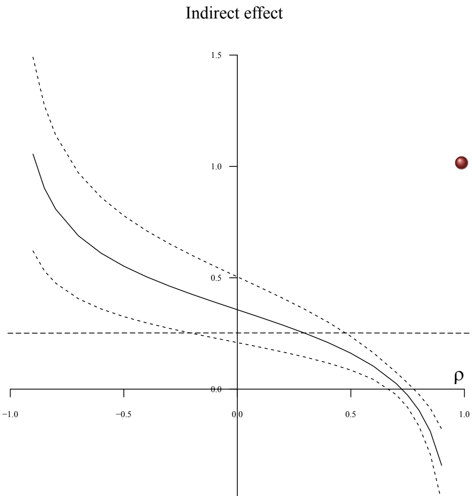


Goal of Sensitivity Analysis

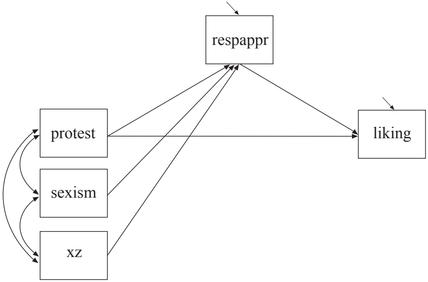
- The residual correlation (ρ) cannot be identified
  - But it can be fixed at different values to see how e.g. the indirect effect changes
- Graph shows effect and its CI as a function of ρ
  - Is the estimated effect still significant for a realistic range of ρ values?

Sensitivity Analysis of Indirect Effect in Simulated Data

- The true indirect effect is 0.25 and is marked by a broken horizontal line
- The standard assumption of $\rho = 0$ mis-estimates the indirect effect as 0.36
- The true ρ is 0.30 which is the x-axis value that gives the true indirect effect



- Conclusions from the graph:
  - The unknown ρ needs to be higher than 0.6 for the effect to be insignificant
  - Such a high ρ value is unlikely: The effect can be considered robust/trustworthy



A moderated mediation model of sex discrimination in the work place. The interaction variable *xz* is the product of the exposure variable protest and the moderator variable sexism ($n = 129$)

- Variables:
  - Protest: binary exposure variable (2 randomized scenarios of female attorney taking action or not)
  - Sexism: Moderator variable
  - Respappr: Mediator - perceived appropriateness of response)
  - Liking: Outcome - how well the subject likes the female attorney

Results for Combined Moderated Mediation for Sex Discrimination

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
liking ON				
respappr	0.098	0.533	0.184	0.854
protest	-3.119	1.750	-1.782	0.075
sexism	-0.462	0.502	-0.919	0.358
mx	0.112	0.157	0.715	0.475
mz	0.039	0.100	0.392	0.695
xz	0.500	0.341	1.466	0.143
respappr ON				
protest	-2.687	1.738	-1.546	0.122
sexism	-0.529	0.320	-1.654	0.098
xz	0.810	0.346	2.343	0.019
Intercepts				
liking	6.510	2.623	2.482	0.013
respappr	6.567	1.596	4.114	0.000
Residual Variances				
liking	0.779	0.135	5.767	0.000
respappr	1.269	0.156	8.121	0.000

Figure: Loop plot of indirect effect and confidence interval for combined moderated mediation case of sex discrimination. The moderator is labeled z in MODEL CONSTRAINT and corresponds to the sexism variable

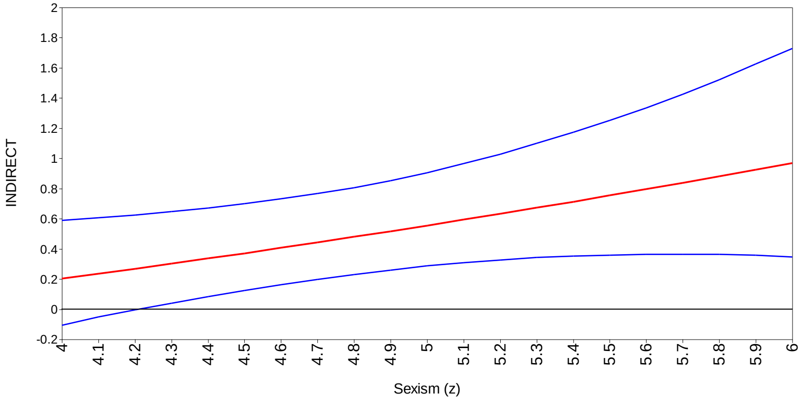
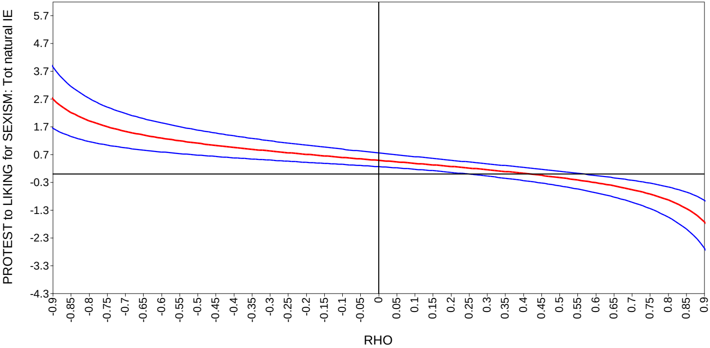


Table: Input for moderated mediation for sex discrimination data

TITLE:	Hayes PROTEST moderation of X ->M, X->Y
DATA:	FILE = protest.txt;
VARIABLE:	NAMES = sexism liking respappr protest;
	USEVARIABLES = liking respappr protest sexism xz;
DEFINE:	xz = protest*sexism;
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 1000;
MODEL:	liking ON respappr (beta1)
	protest (beta2)
	sexism
	xz (beta4);
	respappr ON protest (gamma1)
	sexism (gamma2)
	xz (gamma3);
MODEL INDIRECT:	
	liking MOD respappr sexism(4,6,.1) xz protest;
OUTPUT:	SAMPSTAT STANDARDIZED
	CINTERVAL(BOOTSTRAP);
PLOT:	**TYPE = PLOT3 SENSITIVITY;**

Figure: Sensitivity plot for the indirect effect and its confidence interval at the sexism mean of 5 in a study of sex discrimination in the workplace. The x-axis represents the residual correlation ρ and the y-axis represents the indirect effect



Outline

- Introductory topics
- Count dependent variable
- Censored dependent variable
- Classic mediation analysis
- Sensitivity analysis
- **Modern mediation analysis**
- Bayesian analysis
- Missing data analysis

Mediation Analysis: Modern Counterfactually-Defined Causal Effects

- Why do we need them?
  - The usual $a*b$ indirect effect does not generalize to all models
- Have we used $a*b$ incorrectly?
  - Typically not, but in some cases yes
- Counterfactuals provide a general approach
  - Same effects in many cases: linear models with continuous M and Y
  - New effects e.g. for:
    - Binary M and/or Y
    - Count Y
    - Censored Y
    - Two-part Y

Counterfactually-Defined Causal Effects: Robins, Pearl, VanderWeele, Imai

- Counterfactuals and potential outcomes:
  - Chapter 4: continuous mediator and continuous outcome
  - Chapter 8: continuous mediator and binary outcome, binary mediator and continuous or binary outcome, count outcome, two-part outcome
- Counterfactually-defined causal indirect and direct effects:
  - How are counterfactual effects defined and interpreted?
    - Explanations in pictures, words and formulas
    - Focus on a randomized treatment (1 Tx, 0 Ctrl)

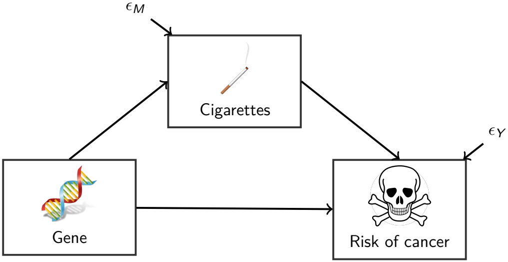
Potential Outcomes, Counterfactuals, and Causal Effects

		Potential Outcomes		Causal effect
i	X_i	Y_i (X_i=1)	Y_i (X_i=0)	Y_i (X_i=1) - Y_i (X_i=0)
1	1	$\boxed{11}$	9	2
2	1	$\boxed{14}$	10	4
3	0	8	$\boxed{5}$	3
4	1	$\boxed{9}$	8	1
5	0	18	$\boxed{12}$	6
6	0	11	$\boxed{10}$	1
True average		11.83	9	2.83
Observed average		11.33	9	2.33

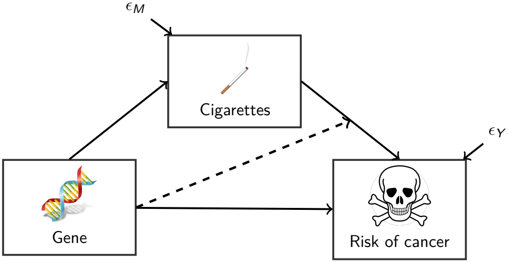
Counterfactually-Defined Causal Effects: Adding a Binary Mediator. Potential Outcomes for M and Y

i	X	M(X=1)	M(X=0)	Y(X=1, M=1)	Y(X=0, M=1)	Y(X=1, M=0)	Y(X=0, M=0)
1	1	$\boxed{1}$	0	$\boxed{11}$			
2	1	$\boxed{1}$	1	$\boxed{14}$			
3	0	0	$\boxed{0}$				$\boxed{5}$
4	1	$\boxed{0}$	0			$\boxed{9}$	
5	0	0	$\boxed{0}$				$\boxed{12}$
6	0	1	$\boxed{1}$		$\boxed{10}$		
Avg		0.667	0.333	12.5	10	9	8.5

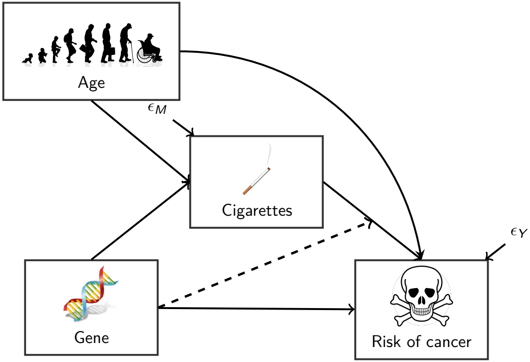
Mediation analysis



Mediation analysis



Mediation analysis

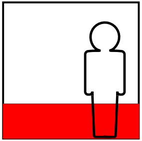


$$Cigarettes_i = \gamma_0 + \gamma_1 \text{Gene}_i + \gamma_2 \text{Age}_i + \epsilon_{Mi}$$

$$Risk_i = \beta_0 + \beta_1 \text{Cigarettes}_i + \beta_2 \text{Gene}_i + \beta_3 \text{Gene}_i \text{Cigarettes}_i + \beta_4 \text{Age}_i + \epsilon_{Yi}$$

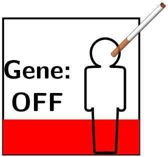
Counterfactually-based causal effects

Expected cancer risk of a person



Counterfactually-based causal effects

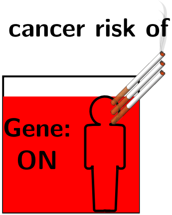
Expected cancer risk of a person



Given not having the gene, smoking as much as non-gene carrier does

Counterfactually-based causal effects

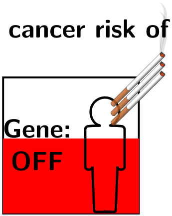
Expected cancer risk of a person



Given having the gene, smoking as much as gene carrier does

Counterfactually-based causal effects

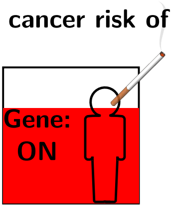
Expected cancer risk of a person



Given not having the gene, smoking as much as gene carrier does

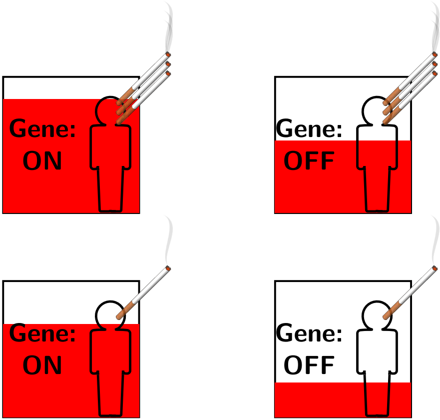
Counterfactually-based causal effects

Expected cancer risk of a person

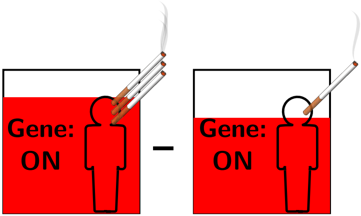


Given having the gene, smoking as much as non-gene carrier does

Counterfactually-based causal effects

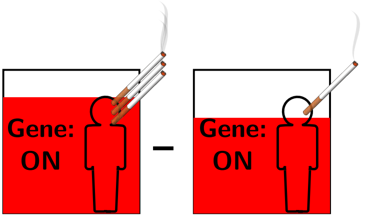


Counterfactually-based causal effects

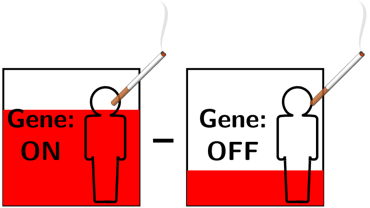
Natural Indirect effect = 

Counterfactually-based causal effects

Natural Indirect effect =



Natural Direct effect =



Counterfactually-based causal effects

Natural Indirect effect $= E[Y(1, M(1))] - E[Y(1, M(0))]$

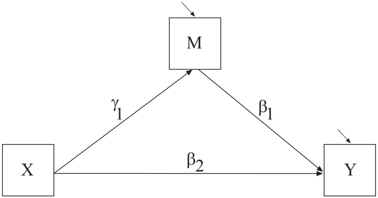
Natural Direct effect $= E[Y(1, M(0))] - E[Y(0, M(0))]$

Counterfactually-based causal effects

Natural Indirect effect $= E[Y(1, M(1))] - E[Y(1, M(0))]$

No functional form is assumed!

Natural Direct effect $= E[Y(1, M(0))] - E[Y(0, M(0))]$



$$Y_i = \beta_0 + \beta_1\, M_i + \beta_2\, X_i + \varepsilon_{yi}$$
$$M_i = \gamma_0 + \gamma_1\, X_i + \varepsilon_{mi}$$

$Y(x_1, M(x_0))$: The variable Y when $X = x_1$ and the variable M varies as it naturally would when $X = x_0$

- Total effect: $E[Y(1,M(1))] - E[Y(0,M(0))]$, treatment group mean of Y minus control group mean of Y
- Total natural indirect effect: $E[Y(1,M(1))] - E[Y(1,M(0))]$
  - $E[Y(1,M(1))]$ is the mean of Y when subjects get the treatment ($X = 1$) and M varies as it would under the treatment condition ($X = 1$) - this is the treatment group mean
  - $E[Y(1,M(0))]$ is the mean of Y when subjects get the treatment ($X = 1$) but M varies as it would under the control condition ($X = 0$) - this is a counterfactual (blocking off effect on Y via M)

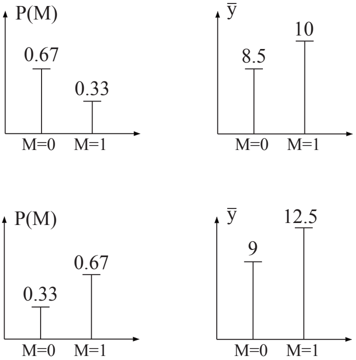
Hypothetical Example with Binary M: Computing Effects

i	X	M(X=1)	M(X=0)	Y(X=1, M=1)	Y(X=0, M=1)	Y(X=1, M=0)	Y(X=0, M=0)
1	1	$\boxed{1}$	0	$\boxed{11}$			
2	1	$\boxed{1}$	1	$\boxed{14}$			
3	0	0	$\boxed{0}$				$\boxed{5}$
4	1	$\boxed{0}$	0			$\boxed{9}$	
5	0	0	$\boxed{0}$				$\boxed{12}$
6	0	1	$\boxed{1}$		$\boxed{10}$		
Avg		0.667	0.333	12.5	10	9	8.5

- Assume that
  - M=1 is more desirable than M=0
  - A high value on Y is more desirable than a low value
- The data on the right show that
  - X=1 (tx) increases the probability of M=1 relative to X=0 (ctrl)
  - M=1 increases the mean of Y relative to M=0

Ctrl, X = 0 :

Tx, X = 1:



$$\hat{E}[Y(0,M(0))] = 8.5*0.67 + 10*0.33 = 9.00\ (ctrl\ grp\ y\ mean)$$
$$\hat{E}[Y(0,M(1))] = 8.5*0.33 + 10*0.67 = 9.50\ (counterfactual)$$
$$\hat{E}[Y(1,M(0))] = 9*0.67 + 12.5*0.33 = 10.17\ (counterfactual)$$
$$\hat{E}[Y(1,M(1))] = 9*0.33 + 12.5*0.67 = 11.33\ (tx\ grp\ y\ mean)$$

The Counterfactual Effects for the Hypothetical Example

- T = total, E = effect, N = natural, I = indirect, D = direct, P = pure

$$TE : \hat{E}[Y(1,M(1))] - \hat{E}[Y(0,M(0))] = 11.33 - 9.00 = 2.33$$
$$TNIE : \hat{E}[Y(1,M(1))] - \hat{E}[Y(1,M(0))] = 11.33 - 10.17 = 1.17$$
$$PNDE : \hat{E}[Y(1,M(0))] - \hat{E}[Y(0,M(0))] = 10.17 - 9.00 = 1.17$$
$$PNIE : \hat{E}[Y(0,M(1))] - \hat{E}[Y(0,M(0))] = 9.50 - 9.00 = 0.50$$
$$TE = TNIE + PNDE$$

- The relationship between M and Y may be different for the two X values, which would indicate an interaction between M and X in their influence on Y
  - This creates a difference between *TNIE* and *PNIE*
  - In this example it is 0.67, that is, the exposure-mediator interaction contributes more than half of the total natural indirect effect *TNIE*
- In this example, estimation is non-parametric, not assuming linear or logistic regression

Indirect Effect $TNIE = E[Y(1,M(1))] - E[Y(1,M(0))]$ in Formulas for a Continuous M

- To get an effect of X on Y we need to integrate out M (collapsing)
- M has two different distributions $f(M|X)$: $M(0)$ for $X = 0$ and $M(1)$ for $X = 1$. For example:
- $E[Y(1,M(0))] = \int_{-\infty}^{+\infty} E[Y|X = 1, M = m] \times f(M|X = 0)\ \partial M$
- In some cases, this integral is simple - integration does not need to be involved: (1) Continuous M, continuous Y, (2) Continuous M, binary Y with probit
- In some cases, the integration is needed: (1) Continuous M, binary Y with logistic (numerical integration needed), (2) Count Y, (3) $log(Y)$

- Continuous M and Y:

$$Y_i = \beta_0 + \beta_1 M_i + \beta_2 X_i + \varepsilon_{yi}, \tag{34}$$
$$M_i = \gamma_0 + \gamma_1 X_i + \varepsilon_{mi}. \tag{35}$$

1. Inserting (35) in (34) and integrating over M,

$$\begin{aligned} E[Y(x_1, M(x_0))] &= \beta_0 + \beta_2 x_1 + \\ &+ \beta_1 \underbrace{\boxed{\int_{-\infty}^{+\infty} M f(M; \gamma_0 + \gamma_1 x_0, \sigma^2)\, \partial M,}}_{E(M \mid X = x_0)} \\ &= \beta_0 + \beta_2 x_1 + \beta_1(\gamma_0 + \gamma_1 x_0). \end{aligned} \tag{36}$$

2. Integration is a general approach but not needed here: Conditioning on $X = x_1$ in (34) and $X = x_0$ in (35) and inserting the mediator expression in the outcome expression, we get the same result:

$$= \beta_0 + \beta_2 x_1 + \beta_1(\gamma_0 + \gamma_1 x_0). \tag{37}$$

Indirect Effect $TNIE = E[Y(x_1, M(x_1))] - E[Y(x_1, M(x_0))]$

- $TNIE$ for continuous M and Y:

$$E[Y(x_1, M(x_1))] - E[Y(x_1, M(x_0))] \tag{33}$$
$$= \beta_0 + \beta_2\, x_1 + \beta_1(\gamma_0 + \gamma_1\, x_1) \tag{34}$$
$$- (\beta_0 + \beta_2\, x_1 + \beta_1(\gamma_0 + \gamma_1\, x_0)) \tag{35}$$
$$= \beta_1\, \gamma_1(x_1 - x_0). \tag{36}$$

- Note 1: Often $x_1 - x_0 = 1$ such as with a one-unit change or treatment/control.
- Note 2: β_0, γ_0, β_2 cancel out. The indirect effect is a product of 2 slopes. This is not the case for binary Y

$$Y_i^* = \beta_0 + \beta_1\, M_i + \beta_2\, X_i + \varepsilon_{yi}, \tag{37}$$

$$M_i = \gamma_0 + \gamma_1\, X_i + \varepsilon_{mi}. \tag{38}$$

Conditioning on $X = x_1$ and $X = x_0$, for Y^* and M, respectively, and inserting M into Y,

$$E(Y^*|X) = \beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_0 + \beta_2\, x_1, \tag{39}$$

$$V(Y^*|X) = V(\beta_1\, \varepsilon_m + \varepsilon_y) = \beta_1^2\, \sigma_m^2 + c. \tag{40}$$

$$P(Y = 1|X) = \Phi[E(Y^*|X)/\sqrt{V(Y^*|X)}], \tag{41}$$

$$TNIE = \Phi[1,1] - \Phi[1,0], \tag{42}$$

where $\Phi[1,1]$ uses $\beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_1 + \beta_2\, x_1$ in $E(Y^*|X)$ and $\Phi[1,0]$ uses $\beta_0 + \beta_1\, \gamma_0 + \beta_1\, \gamma_1\, x_0 + \beta_2\, x_1$. All 6 parameters involved.

Effects Expressed on an Odds Ratio Scale for a Binary Outcome: Probit Model

The total natural indirect effect odds ratio for a binary exposure can be expressed as

$$TNIE(OR) = \frac{P(Y_{x_1 M_{x_1}} = 1)/(1 - P(Y_{x_1 M_{x_1}} = 1)}{P(Y_{x_1 M_{x_0}} = 1)/(1 - P(Y_{x_1 M_{x_0}} = 1))}$$
$$= \frac{\Phi[probit(1,1)]/(1 - \Phi[probit(1,1)])}{\Phi[probit(1,0)]/(1 - \Phi[probit(1,0)])}. \qquad (43)$$

Odds Ratio Effects Assuming a Rare Binary Outcome: Logistic Model

VanderWeele and Vansteelandt (2010) show that with logistic regression the TNIE odds ratio is approximately equal to

$$TNIE(OR) \approx e^{\beta_1\, \gamma_1 + \beta_3\, \gamma_1}, \tag{44}$$

that is, the indirect effect odds ratio uses the same formula as the indirect effect with a continuous outcome, but exponentiated. When the treatment variable is continuous, the indirect effect odds ratio of (44) is modified as

$$TNIE(OR) = e^{(\beta_1\, \gamma_1 + \beta_3\, \gamma_1\, x_1)(x_1 - x_0)}, \tag{45}$$

for a change from x_0 to x_1. For example, x_0 may represent the mean of the treatment and x_1 may represent the mean plus one standard deviation, so that $x_1 - x_0$ corresponds to one standard deviation for the continuous treatment variable.

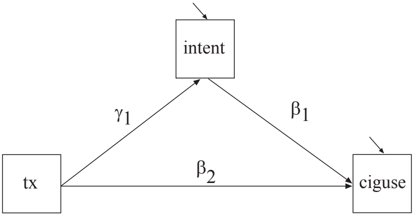
Valeri and VanderWeele (2013):

> *"In summary, controlled direct effects require (a) no unmeasured treatment-outcome confounding and (b) no unmeasured mediator-outcome confounding. Natural direct and indirect effects require these assumptions and also (c) no unmeasured treatment-mediator confounding and (d) no mediator-outcome confounder affected by treatment. It is important to note that randomizing the treatment is not enough to rule out confounding issues in mediation analysis. This is because randomization of the treatment rules out the problem of treatment-outcome and treatment-mediator confounding but does not guarantee that the assumption of no confounding of mediator-outcome relationship holds. This is because even if the treatment is randomized, the mediator generally will not be."*

Muthén et al. (2016):

> *"One may certainly question if effects in mediation analysis can be considered more causal with the advent of counterfactually-defined effects. On a positive note, however, one can claim that it is better to use effects that under some well-defined circumstances are causal even if in a particular application one is not sure that the assumptions are fulfilled. The strength of the counterfactual approach is that it provides a road map for how to go about defining the effects in the first place."*

Example: Smoking Data



Drug intervention program for students in Grade 6 and Grade 7 in Kansas City schools ($n = 864$). MacKinnon et al. (2007), Clinical Trials.

- Schools were randomly assigned to the treatment or control group (the multilevel aspect of the data is ignored)
- The mediator is the intention to use cigarettes in the following 2-month period which was measured about six months after baseline
- The outcome is cigarette use or not in the previous month which was measured at follow-up
- Cigarette use is observed for 18% of the sample

- The total effect can be computed without doing a mediation analysis as the difference between the proportion of smokers in the treatment group and the proportion of smokers in the control group
- This results in an estimate of the total effect as the difference in the probabilities of $0.148 - 0.224 = -0.076$
- The corresponding estimate of the total effect odds ratio is

$$TE(OR) = \frac{0.148/(1-0.148)}{0.224/(1-0.224)} = 0.602. \tag{46}$$

- Both estimates indicate a lowering of the smoking probability due to treatment

Table: Input for smoking data using probit

TITLE:	Clinical Trials data from MacKinnon et al. (2007)
DATA:	FILE = smoking.txt;
VARIABLE:	NAMES = intent tx ciguse;
	USEVARIABLES = tx ciguse intent;
	CATEGORICAL = ciguse;
ANALYSIS:	**ESTIMATOR = ML;**
	LINK = PROBIT;
	BOOTSTRAP = 10000;
MODEL:	ciguse ON intent tx;
	intent ON tx;
MODEL INDIRECT:	
	ciguse IND intent tx;
OUTPUT:	TECH1 TECH8 SAMPSTAT
	CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Table: Bootstrap confidence intervals for smoking data effects using probit regression for the outcome cigarette

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
Effects from tx to ciguse					
Tot natural IE	-0.040	-0.036	-0.022	-0.008	-0.006
Pure natural DE	-0.104	-0.095	-0.050	-0.005	0.004
Total effect	-0.128	-0.119	-0.072	-0.026	-0.017
Odds ratios for binary Y					
Tot natural IE	0.757	0.772	0.853	0.939	0.958
Pure natural DE	0.520	0.551	0.731	0.969	1.025
Total effect	0.433	0.461	0.624	0.841	0.896

Effects for Smoking Data Using Probit

- The total natural indirect effect (TNIE) in probability metric is estimated as -0.022 and is significant because the 95% confidence interval does not cover zero: $[-0.040, -0.006]$
- The indirect effect odds ratio is estimated as 0.853 and is significant because the 95% confidence interval does not cover one: $[0.757, 0.958]$
- The direct effect in probability metric is estimated as -0.050 and is not significant. The direct effect odds ratio of 0.731 is not significant
- The total effect in probability metric of -0.072 is significant
- The total effect can be compared to the proportion of cigarette users in the control group of 0.224. This shows a drop of 32% due to treatment (0.072/0.224 = 0.32)

Table: Input for smoking data using logistic regression for the cigarette use outcome

TITLE:	Clinical Trials data from MacKinnon et al. (2007)
DATA:	FILE = smoking.txt;
VARIABLE:	NAMES = intent tx ciguse;
	USEVARIABLES = tx ciguse intent;
	CATEGORICAL = ciguse;
ANALYSIS:	ESTIMATOR = ML;
	LINK = LOGIT;
	BOOTSTRAP = 10000;
MODEL:	ciguse ON intent (beta1)
	tx (beta2);
	intent ON tx (gamma);
MODEL INDIRECT:	
	ciguse IND intent tx;
MODEL CONSTRAINT:	
	NEW(indirect direct);
	indirect = EXP(beta1*gamma);
	direct = EXP(beta2);
OUTPUT:	TECH1 TECH8 SAMPSTAT
	CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

- Not assuming a rare outcome (using MODEL INDIRECT): TNIE (OR) $= 0.858$, TNDE (OR) $= 0.716$
- Assuming a rare outcome (using MODEL CONSTRAINT): TNIE (OR) $= 0.843$, TNDE (OR) $= 0.686$

- The rare outcome results indicate stronger effects with estimates farther from one
- The rare outcome assumption may not be suitable here with 18% smoking prevalence

- Probit and logistic give similar results

Moderated Mediation with a Binary Outcome: Vaccination

- Hopfer (2012) analyzed data from a randomized control trial aimed at increasing the vaccination rate for the human papillomavirus (HPV) among college women ($n = 394$)
  - Subjects were randomized into three different intervention groups and a control group where the groups were presented with different forms of video with vaccine decision narratives
  - The mediator measures intent to get vaccinated
  - Control variables are HPV communication with parents (yes/no), age, sexually active (yes/no), and HPV knowledge
  - Only the effects of the combined peer-expert intervention are considered ($tx2$)
  - In this group, to which 25% of the sample was randomized, the vaccination rate is 22.2% whereas in the control group it is 12.0%
  - This gives an estimate of the total intervention effect in the probability metric of 0.10 and in the odds ratio metric of 2.70

Figure: Moderated mediation model for the HPV vaccination data using a logistic regression for the vaccination outcome

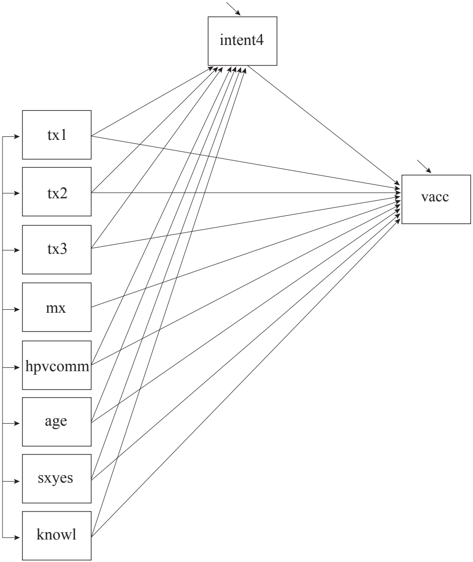


Table: Input for the model with intervention-mediator interaction for HPV vaccination data

VARIABLE:	
	USEVARIABLES = intent4 tx1 tx2 tx3 vacc hpvcomm age sxyes knowl mx;
	CATEGORICAL = vacc;
	MISSING = ALL (99);
DEFINE:	**mx = intent4*tx2;**
	CENTER age knowl(GRANDMEAN);
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
MODEL:	vacc ON intent4 tx1 tx2 tx3 hpvcomm age sxyes knowl mx;
	intent4 ON tx1 tx2 tx3 hpvcomm age sxyes knowl;
MODEL INDIRECT:	
	vacc MOD intent4 mx tx2;
OUTPUT:	SAMPSTAT PATTERNS CINTERVAL(BOOTSTRAP) TECH1 TECH8;
PLOT:	TYPE = PLOT3;

Table: Results for HPV vaccination data

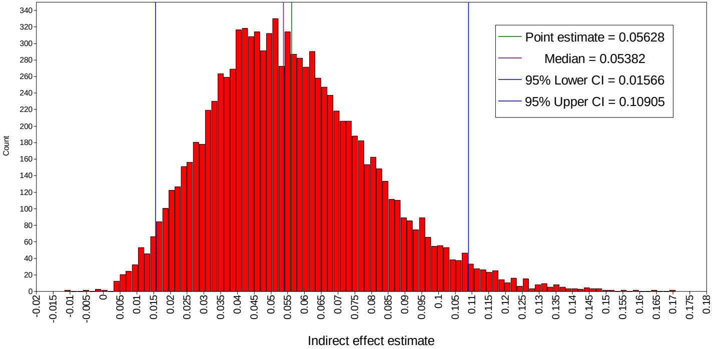
	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
vacc ON				
intent4	1.303	0.262	4.974	0.000
tx1	0.320	0.435	0.735	0.463
tx2	-1.180	2.271	-0.520	0.603
tx3	-0.818	2.141	-0.382	0.703
hpvcomm	0.242	0.350	0.693	0.488
age	0.194	0.084	2.311	0.021
sxyes	0.219	0.333	0.658	0.511
knowl	-0.041	0.072	-0.572	0.568
mx	0.494	0.660	0.749	0.454
intent4 ON				
tx1	0.149	0.106	1.400	0.161
tx2	0.300	0.092	3.270	0.001
tx3	-0.066	0.141	-0.465	0.642
hpvcomm	0.093	0.078	1.196	0.232
age	-0.049	0.021	-2.283	0.022
sxyes	0.044	0.078	0.573	0.567
knowl	-0.003	0.017	-0.160	0.873
Intercepts				
intent4	2.718	0.082	32.959	0.000
Thresholds				
vacc$1	6.227	0.877	7.100	0.000
Residual Variances				
intent4	0.591	0.041	14.293	0.000

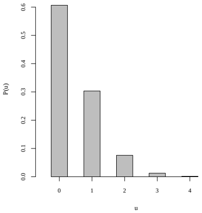
Table: Bootstrap confidence intervals without and with intervention-mediator interaction for HPV vaccination data

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)

	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
Without intervention-mediator interaction					
Effects from TX2 to VACC					
Tot natural IE	0.016	0.020	0.048	0.083	0.092
Pure natural DE	-0.019	-0.010	0.041	0.098	0.111
Total effect	0.013	0.024	0.089	0.165	0.182
Odds ratios for binary Y					
Tot natural IE	1.155	1.197	1.448	1.833	1.932
Pure natural DE	0.803	0.894	1.523	2.715	3.045
Total effect	1.137	1.283	2.205	4.115	4.665
With intervention-mediator interaction					
Effects from TX2 to VACC					
Tot natural IE	0.016	0.020	0.056	0.099	0.109
Pure natural DE	-0.022	-0.012	0.037	0.095	0.107
Total effect	0.016	0.028	0.093	0.169	0.186
Odds ratios for binary Y					
Tot natural IE	1.147	1.200	1.541	2.096	2.238
Pure natural DE	0.773	0.865	1.467	2.662	2.964
Total effect	1.178	1.313	2.260	4.234	4.791

Figure: Bootstrap distribution for the total natural indirect effect estimate in probability metric for the model with intervention-mediator interaction for the HPV vaccination data





$$log\mu_i = \beta_0 + \beta_1\, M_i + \beta_2\, X_i + \beta_3\, MX_i + \beta_4\, C_i, \tag{47}$$

$$M_i = \gamma_0 + \gamma_1\, X_i + \gamma_2\, C_i + \varepsilon_{mi}. \tag{48}$$

As before, the counterfactually-based causal effects consider terms such as

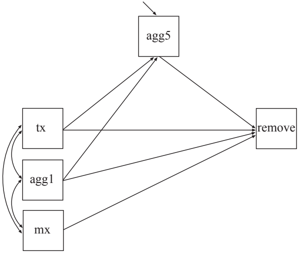
$$E[Y(x_1, M(x_0))] = \int_{-\infty}^{\infty} E[Y \mid C = c, X = x_1, M = m] \tag{49}$$

$$\times f(M \mid C = c, X = x_0)\, \partial M. \tag{50}$$

This needs to take into account that the rate (mean) is

$$E[Y \mid C = c, X = x_1, M = m] = e^{\beta_0 + \beta_1\, m + \beta_2\, x_1 + \beta_3\, m\, x_1 + \beta_4\, c}. \tag{51}$$

Example: A Mediation Model for Aggressive Behavior and a School Removal Count Outcome: Case 3 (*mx*) Moderation



Randomized field experiment in Baltimore public schools with a classroom-based intervention aimed at reducing aggressive-disruptive behavior among elementary school students (Kellam et al., 2008). The analysis uses $n = 250$ boys.

- The outcome variable remove is the number of times a student has been removed from school during grades 1-7
- tx is the binary exposure variable representing the intervention
- The Fall baseline aggression score is agg1 which was observed before the intervention started
- The mediator variable agg5 is the Grade 5 aggression score.
- An intervention-mediator interaction variable *mx* is included to moderate the influence of the mediator on the outcome.

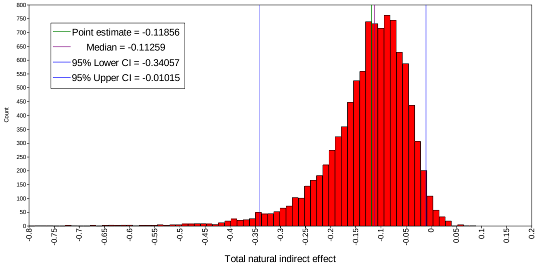
Table: Input for negative binomial model for school removal data

VARIABLE:	
	USEVARIABLES = remove agg5 agg1 tx mx;
	IDVARIABLE = prcid;
	COUNT = remove(NB);
	USEOBSERVATIONS = gender EQ 1 AND (desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3 OR desgn11s EQ 4);
DEFINE:	IF(desgn11s EQ 4)THEN tx=1;
	IF(desgn11s EQ 1 OR desgn11s EQ 2 OR desgn11s EQ 3)THEN tx=0;
	remove = total17;
	agg1 = sctaa11f;
	agg5 = sctaa15s;
	CENTER agg1 agg5(GRANDMEAN);
	mx = agg5*tx;
ANALYSIS:	ESTIMATOR = ML;
	BOOTSTRAP = 10000;
	PROCESSORS = 8;
MODEL:	remove ON agg5 tx mx agg1;
	agg5 ON tx agg1;
MODEL INDIRECT:	
	remove MOD agg5 mx tx;
OUTPUT:	SAMPSTAT TECH1 TECH8 PATTERNS CINTERVAL(BOOTSTRAP);
PLOT:	TYPE = PLOT3;

Table: Bootstrap confidence intervals for effects for school removal data

Confidence intervals of total, indirect, and direct effects based on counterfactuals (causally-defined effects)					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
		Effects from TX to REMOVE			
Tot natural IE	-0.341	-0.283	-0.119	-0.024	-0.010
Pure natural DE	-0.681	-0.608	-0.272	0.125	0.213
Total effect	-0.794	-0.722	-0.391	-0.032	0.034
		Other effects			
Pure natural IE	-0.358	-0.327	-0.183	-0.050	-0.023
Tot natural DE	-0.587	-0.525	-0.208	0.135	0.213
Total effect	-0.794	-0.722	-0.391	-0.032	0.034

Figure: Total natural indirect effect bootstrap distribution for school removal data



- The indirect effect estimate -0.119 is in a log rate metric for the count outcome of school removal and is hard to interpret
  - One way to make the effect size understandable is to compute the probability of a zero count
  - The intervention increases the probability of a zero school removals from 0.294 to 0.435

- Let $U = 1$ refer to the event of not being at the floor of the outcome
- Probit regression is used to describe the probability of $U = 1$
- For those not at the floor value, the outcome is transformed using the natural logarithm to make the normality assumption more realistic

The two-part mediation model with a control variable C and a treatment-mediator interaction MX is expressed as

$$log\ Y_{i|U_i=1} = \beta_0 + \beta_1\, M_i + \beta_2\, X_i + \beta_3\, MX_i + \beta_4\, C_i + \varepsilon_{yi}, \tag{52}$$

$$M_i = \gamma_0 + \gamma_1\, X_i + \gamma_2\, C_i + \varepsilon_{mi}, \tag{53}$$

$$probit(\pi_i) = \kappa_0 + \kappa_1\, M_i + \kappa_2\, X_i + \kappa_3\, MX_i + \kappa_4\, C_i, \tag{54}$$

where the residual $\varepsilon_y \sim N(0, \sigma_y^2)$, the residual $\varepsilon_m \sim N(0, \sigma_m^2)$, and π_i represents the probability of not being at the floor,

- Example from Hayes (2013):
  - $n = 262$ small-business owners' economic stress (Pollack et al., 2011)
  - The exposure variable is a continuous variable representing economic stress
  - The mediator variable is a continuous variable representing depressed affect
  - The outcome variable is a continuous variable representing thoughts about withdrawing from their entrepreneurship

The outcome variable withdraw has a 30% floor effect:

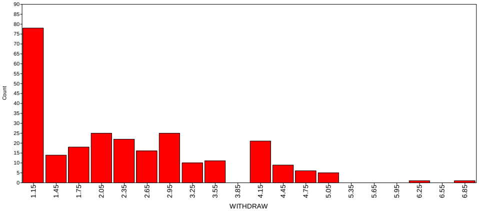


Table: Input for two-part mediation modeling of economic stress data

TITLE:	Hayes ESTRESS example, cont's X
DATA:	FILE = estress.txt;
VARIABLE:	NAMES = tenure estress affect withdraw sex age ese;
	USEVARIABLES = affect estress u y;
	CATEGORICAL = u;
DEFINE:	**withdraw = withdraw - 1;**
DATA TWOPART:	
	NAMES = withdraw;
	BINARY = u;
	CONTINUOUS = y;
	CUTPOINT = 0;
ANALYSIS:	ESTIMATOR = ML;
	LINK = PROBIT;
	BOOTSTRAP = 1000;
MODEL:	y ON affect (beta1)
	estress (beta2);
	[y] (beta0);
	y (v);
	affect ON estress (gamma1);
	[affect] (gamma0);
	affect (sig);
	u ON affect (kappa1)
	estress (kappa2);
	[u$1] (kappa0);
MODEL INDIRECT:	
	u IND affect estress (6.04 4.62);
	-table continues-

Table: Input for two-part mediation modeling of economic stress data

```
MODEL CONSTRAINT:
            NEW(x1 x0 ey1 ey0 mum1 mum0 ay1 ay0 bym11 bym10 bym01
            bym00 eym11 eym10 eym01 eym00 tnie pnde total pnie beta3 sd pi11
            pi10 pi01 pi00);
            beta3 = 0;
            x1=6.04;
            x0=4.62;
            ey1=EXP(v/2)*EXP(beta0+beta2*x1);
            ey0=EXP(v/2)*EXP(beta0+beta2*x0);
            mum1=gamma0+gamma1*x1;
            mum0=gamma0+gamma1*x0;
            ay1=sig*(beta1+beta3*x1);
            ay0=sig*(beta1+beta3*x0);
            bym11=(ay1/mum1+1);
            bym10=(ay1/mum0+1);
            bym01=(ay0/mum1+1);
            bym00=(ay0/mum0+1);
            sd=SQRT(kappa1*kappa1*sig+1);
            pi11=PHI((-kappa0+kappa2*x1+kappa1*bym11*
            (gamma0+gamma1*x1))/sd);
            pi10=PHI((-kappa0+kappa2*x1+kappa1*bym10*
            (gamma0+gamma1*x0))/sd);
            pi01=PHI((-kappa0+kappa2*x0+kappa1*bym11*
            (gamma0+gamma1*x1))/sd);
            pi00=PHI((-kappa0+kappa2*x0+kappa1*bym00*
            (gamma0+gamma1*x0))/sd);
            eym11=EXP((bym11*bym11-1)*mum1*mum1/(2*sig));
            eym10=EXP((bym10*bym10-1)*mum0*mum0/(2*sig));
            eym01=EXP((bym01*bym01-1)*mum1*mum1/(2*sig));
            eym00=EXP((bym00*bym00-1)*mum0*mum0/(2*sig));
            tnie=pi11*ey1*eym11-pi10*ey1*eym10;
            pnde=pi10*ey1*eym10-pi00*ey0*eym00;
            total=pi11*ey1*eym11-pi00*ey0*eym00;
            pnie=pi01*ey0*eym01-pi00*ey0*eym00;
PLOT:       TYPE = PLOT3;
OUTPUT:     SAMPSTAT TECH1 TECH8
            CINTERVAL(BOOTSTRAP);
```

Table: Bootstrap confidence intervals for four mediation models

Confidence intervals for effects					
	Lower 2.5%	Lower 5%	Estimate	Upper 5%	Upper 2.5%
(1) Two-part: overall effects for the outcome					
TNIE	0.104	0.121	0.203	0.293	0.311
PNDE	-0.304	-0.276	-0.145	-0.011	0.019
TE	-0.124	-0.089	0.058	0.207	0.246
(2) Two-part: effects for binary part of the outcome					
TNIE	0.036	0.041	0.071	0.103	0.108
PNDE	-0.074	-0.062	-0.016	0.028	0.035
TE	-0.006	0.005	0.055	0.098	0.105
(3) Two-part: conditional effects for continuous part of the outcome					
TNIE	0.043	0.053	0.112	0.177	0.194
PNDE	-0.322	-0.299	-0.160	-0.008	0.023
TE	-0.219	-0.184	-0.048	0.105	0.131
(4) Regular: effects using log y					
TNIE	0.098	0.108	0.182	0.267	0.284
PNDE	-0.236	-0.209	-0.084	0.044	0.066
TE	-0.072	-0.045	0.099	0.243	0.269
(5) Regular: effects using the original y					
TNIE	0.103	0.117	0.189	0.266	0.282
PNDE	-0.263	-0.243	-0.109	0.027	0.051
TE	-0.116	-0.069	0.080	0.220	0.245

Further Mediation Topics

- The book also covers:
  - Ordinal M, Y
  - Nominal M
- Binary outcome with multiple mediators
  - Nguyen et al. (2016). Causal mediation analysis with a binary outcome and multiple continuous or ordinal mediators: Simulations and application to an alcohol intervention. Structural Equation Modeling: A Multidisciplinary Journal, 23:3, 368-383
- Counterfactually-defined, path-specific effects with multiple mediators
  - Steen et al. (2017). Flexible mediation analysis with multiple mediators. American Journal of Epidemiology, 186, 184-193.

Further Mediation Topics Continued

- Longitudinal mediation
  - Maxwell & Cole (2007). Bias in cross-sectional analyses of longitudinal mediation. Psychological Methods
  - Deboeck & Preacher (2015). No need to be discrete: A method for continuous time mediation analysis. Structural Equation Modeling
  - Vanderweele & Tchetgen (2017). Mediator analysis with time varying exposures and mediators. Journal of the Royal Statistical Society, Series B, 79, 917-938
- Multilevel mediation
  - Preacher, Zhang, & Zyphur (2011) Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. Structural Equation Modeling, 18, 161-182

Outline

- Introductory topics
- Count dependent variable
- Censored dependent variable
- Classic mediation analysis
- Sensitivity analysis
- Modern mediation analysis
- **Bayesian analysis**
- Missing data analysis

All that's needed:

ESTIMATOR = BAYES;

- Advantages over ML
- Prior, likelihood, posterior
- Iterations, convergence, plots, model fit
- Mediation examples: non-informative and informative priors

Bayesian Analysis: Advantages over ML

- Six key advantages of Bayesian analysis over frequentist analysis using maximum likelihood estimation:
  1. More can be learned about parameter estimates and model fit
  2. Large-sample theory is not needed and small-sample performance is better
  3. Parameter priors can better reflect results of previous studies
  4. Analyses are in some cases less computationally demanding, for example, when maximum-likelihood requires high-dimensional numerical integration
  5. In cases where maximum-likelihood computations are prohibitive, Bayes with non-informative priors can be viewed as a computing algorithm that would give essentially the same results as maximum-likelihood if maximum-likelihood estimation were computationally feasible
  6. New types of models can be analyzed where the maximum-likelihood approach is not practical (e.g. DSEM)

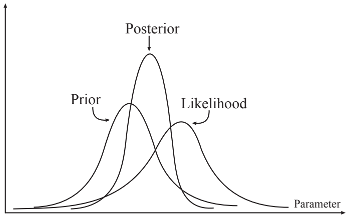


Figure: Informative prior

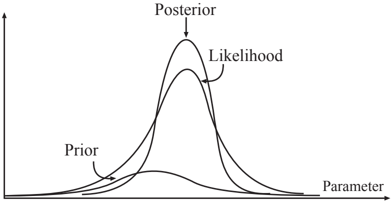
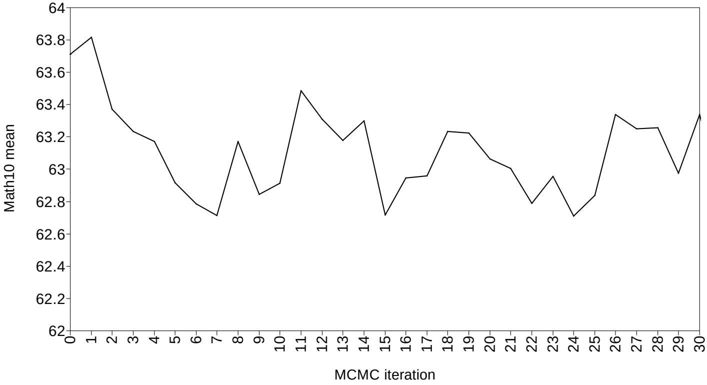


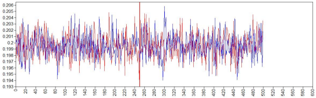
Figure: Non-informative prior

- Priors:
  - Non-informative priors (diffuse priors): Large variance (default in Mplus)
    - A large variance reflects large uncertainty in the parameter value. As the prior variance increases, the Bayesian estimate gets closer to the maximum-likelihood estimate
  - Weakly informative priors: Used for technical assistance
  - Informative priors:
    - Informative priors reflect prior beliefs in likely parameter values
    - These beliefs may come from substantive theory combined with previous studies of similar populations

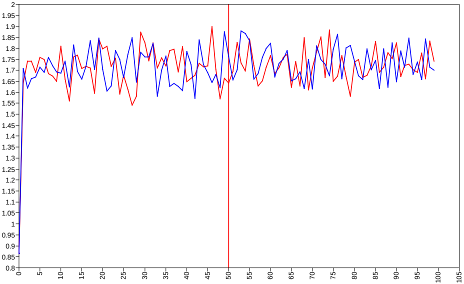
Trace Plot: 20 MCMC Iterations for the LSAY Math10 Mean



- Starting value for the mean is the listwise estimate of 63.7


0.206
0.205
0.204
0.203
0.202
0.201
0.2
0.199
0.198
0.197
0.196
0.195
0.194
0.193
0 20 40 60 80 100 120 140 160 180 200 220 240 260 280 300 320 340 360 380 400 420 440 460 480 500 520 540 560 580 600

Convergence: Trace Plot for Two MCMC Chains. PSR

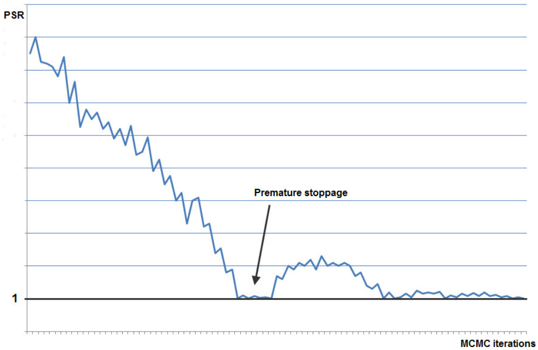


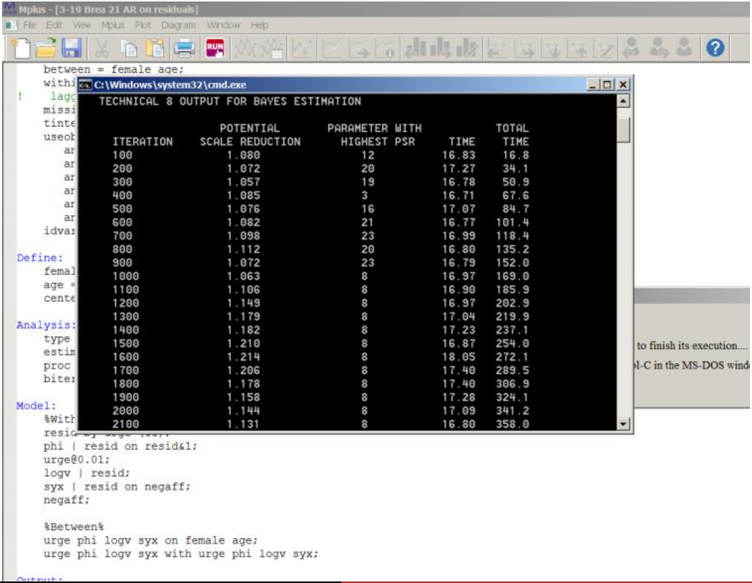
Potential scale reduction criterion (Gelman & Rubin, 1992):

$$PSR = \sqrt{\frac{W+B}{W}}, \quad (55)$$

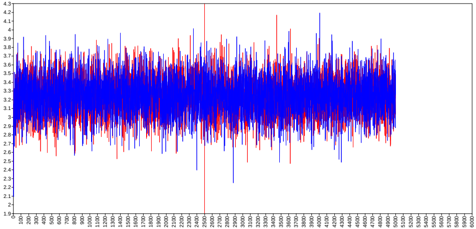
where W represents the within-chain variation of a parameter and B represents the between-chain variation of a parameter. **A PSR value close to 1 means that the between-chain variation is small relative to the within-chain variation** and is considered evidence of convergence.

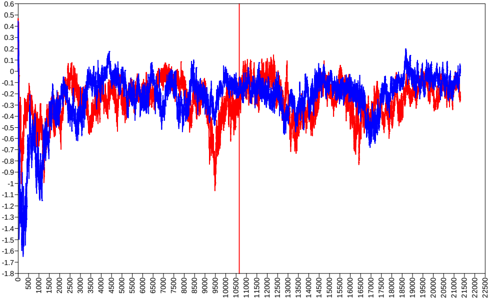
Figure: Premature stoppage of Bayes MCMC iterations using the Potential Scale Reduction (PSR) criterion



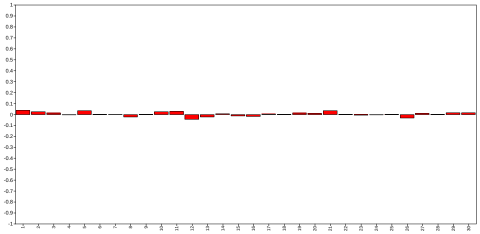

Mplus - [3-19 Brea 21 AR on residuals]
File Edit View Mplus Plot Diagram Window Help
RUN
between = female age;
Define:
Analysis:
Model:
phi | resid on resid&1;
urge@0.01;
logv | resid;
syx | resid on negaff;
negaff;
%Between%
urge phi logv syx on female age;
urge phi logv syx with urge phi logv syx;
C:\Windows\system32\cmd.exe
TECHNICAL 8 OUTPUT FOR BAYES ESTIMATION
ITERATION | POTENTIAL SCALE REDUCTION | PARAMETER WITH HIGHEST PSR | TIME | TOTAL TIME
100 | 1.080 | 12 | 16.83 | 16.8
200 | 1.072 | 20 | 17.27 | 34.1
300 | 1.057 | 19 | 16.78 | 50.9
400 | 1.085 | 3 | 16.71 | 67.6
500 | 1.076 | 16 | 17.07 | 84.7
600 | 1.082 | 21 | 16.77 | 101.4
700 | 1.098 | 23 | 16.99 | 118.4
800 | 1.112 | 20 | 16.80 | 135.2
900 | 1.072 | 23 | 16.79 | 152.0
1000 | 1.063 | 8 | 16.97 | 169.0
1100 | 1.106 | 8 | 16.90 | 185.9
1200 | 1.149 | 8 | 16.97 | 202.9
1300 | 1.179 | 8 | 17.04 | 219.9
1400 | 1.182 | 8 | 17.23 | 237.1
1500 | 1.210 | 8 | 16.87 | 254.0
1600 | 1.214 | 8 | 18.05 | 272.1
1700 | 1.206 | 8 | 17.40 | 289.5
1800 | 1.178 | 8 | 17.40 | 306.9
1900 | 1.158 | 8 | 17.28 | 324.1
2000 | 1.144 | 8 | 17.09 | 341.2
2100 | 1.131 | 8 | 16.80 | 358.0
to finish its execution....

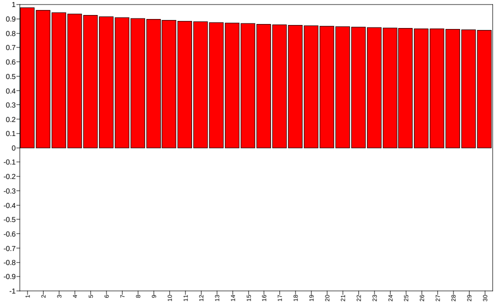
Trace Plots Indicating Good vs Poor Mixing



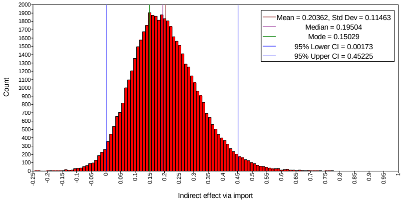


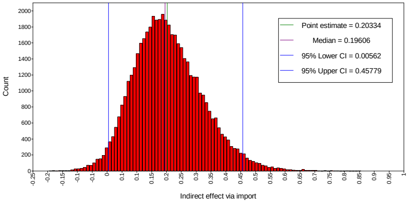
Autocorrelation Plots Indicating Good vs Poor Mixing





Bayes Posterior Distribution Similar to ML Bootstrap Distribution: Credibility versus Confidence Intervals





Speed Of Bayes In Mplus

Wang & Preacher (2014). Moderated mediation analysis using Bayesian methods. *Structural Equation Modeling*.

- Comparison of ML (with bootstrap) and Bayes: Similar statistical performance
- Comparison of Bayes using BUGS versus Mplus: Mplus is 15 times faster

- Reason for Bayes being faster in Mplus:
  - Mplus uses Fortran (fastest computational environment)
  - Mplus uses parallel computing so each chain is computed separately
  - Mplus uses the largest updating blocks possible - complicated to program but gives the best mixing quality
  - Mplus uses sufficient statistics when possible
- Mplus Bayes considerably easier to use

Examples of Bayes' Advantage Over ML

- Non-informative priors: Bayes as a computationally less demanding computing algorithm than ML
- Informative priors: Bayes as a better reflection of substantive theory

Bayes' Advantage Over ML: Non-Informative Priors Missing Data with a Binary Outcome

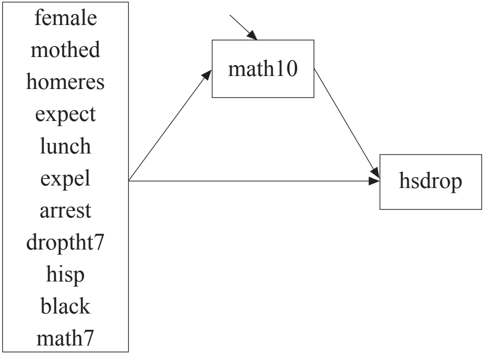
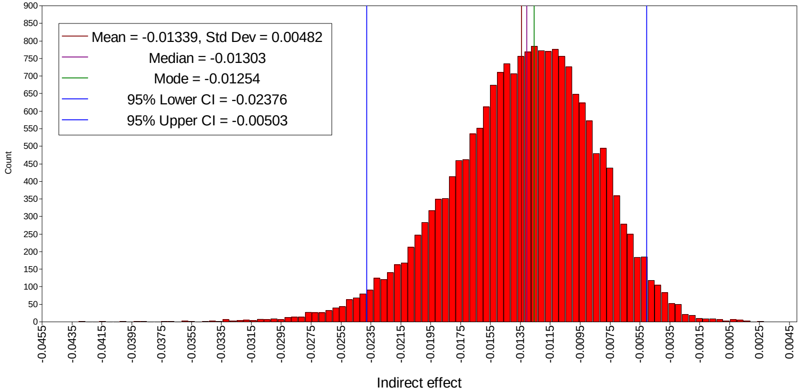


Figure: Mediation model for a binary outcome of dropping out of high school (n=2898)

```
                  CATEGORICAL = hsdrop;
ANALYSIS:         ESTIMATOR = BAYES;
                  PROCESSORS = 2;
                  BITERATIONS = (20000);
MODEL:            hsdrop ON math10 female-math7;
                  math10 ON female-math7;
MODEL INDIRECT:
                  hsdrop IND math10 math7(61.01 50.88);
OUTPUT:           SAMPSTAT PATTERNS TECH1 TECH8 CINTERVAL;
PLOT:             TYPE = PLOT3;
```

Indirect and direct effects computed in probability scale using counterfactually-based causal effects.

Bayesian Posterior Distribution Of Indirect Effect For High School Dropout



ML estimates are almost identical to Bayes, but:

- ML needs Monte Carlo integration with 250 points because the mediator is a partially latent variable due to missing data
- ML needs bootstrapping (1,000 draws) to capture CIs for the non-normal indirect effect
- ML takes 21 minutes

- Bayes takes 21 seconds
- Bayes posterior distribution for the indirect effect is based on 20,000 draws as compared to 1,000 bootstraps for ML

Bayes' Advantage Over ML: Informative Priors In A Mediation Model

- Yuan and MacKinnon (2009) in Psychological Methods
- $n = 354$ firefighters
  - x: exposure to randomized experiment
  - m: change in knowledge of the benefits of healthy eating
  - y: reported healthy eating
- Priors for a and b from previous studies - mean and variance
  - $a \sim N(0.35, 0.04)$
  - $b \sim N(0.1, 0.01)$
  - Prior variances set as 4 times larger than observed to account for study differences
  - The credibility interval for the indirect effect is 16% shorter using the priors
    - With a smaller sample the priors have a larger effect

```
TITLE:            Yuan and MacKinnon firefighters mediation using Bayesian analysis
                  Elliot DL, Goldberg L, Kuehl KS, et al. The PHLAME Study: proces
                  outcomes of 2 models of behavior change.
                  J Occup Environ Med. 2007;49(2):204-213.
DATA:             FILE = fire.dat;
VARIABLE:         NAMES = y m x;
MODEL:            y ON m (b)
                  x;
                  m ON x (a);
ANALYSIS:         ESTIMATOR = BAYES;
                  PROCESSORS = 2;
                  BITERATIONS = (20000);
MODEL PRIORS:
                  a~N(0.35,0.04);
                  b~N(0.1,0.01);
MODEL INDIRECT:
                  y IND x;
OUTPUT:           SAMPSTAT TECH1 TECH8 CINTERVAL;
PLOT:             TYPE = PLOT3;
```

Outline

- Introductory topics
- Count dependent variable
- Censored dependent variable
- Classic mediation analysis
- Sensitivity analysis
- Modern mediation analysis
- Bayesian analysis
- **Missing data analysis**

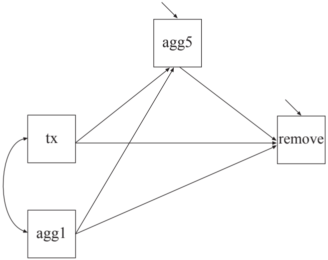
Missing Data

- Missing data descriptives
- MCAR, MAR, and NMAR definitions
- MAR in a simple bivariate case
- Modeling the missing data
- (Multiple imputation)
  - Advantage: Can use variables not in the model
  - Disadvantage: Limited later analysis options
- Auxiliary variables: Making MAR more plausible
- Missing on covariates

Take Home Messages

- Bottom line: Choose ML or Bayes estimation
- ML (or Bayes) under MAR is the default in Mplus: No need to do anything but give the missing data flag
  - Both ML and Bayes use all available data which is optimal
  - ML is sometimes called FIML but is simply ML under the "MAR" assumption
  - Bayes is advantageous with missing on binary covariates

Figure: Mediation model for aggressive behavior in the classroom. The outcome variable remove measures how many times a student was removed from class.



How is the analysis affected if we have missing data on the outcome, the mediator, or the control variables (covariates)?

Missing Data Patterns and Coverage for Aggression Data

Missing data patterns (x = not missing)

	1	2	3	4
tx	x	x	x	x
remove	x	x	x	x
agg1	x	x		
agg5	x		x	

Missing data pattern frequencies

Pattern	Frequency	Pattern	Frequency
1	250	3	21
2	142	4	28

Covariance coverage

	tx	remove	agg1	agg5
tx	1.000			
remove	1.000	1.000		
agg1	0.889	0.889	0.889	
agg5	0.615	0.615	0.567	0.615

Types of Missingness: MCAR, MAR, and NMAR

- MCAR (Missing Completely At Random)
  - Missingness not a function of any observed or latent variable in the model or outside the model
  - If MCAR holds, listwise deletion is ok except for loss of power
  - Typically only achieved if designed: Random forms
- MAR (Missing At Random)
  - Missingness allowed to be a function of the observed variables in the model
  - Standard assumption for ML and Bayes
  - Not possible to test if MAR holds
- NMAR (Not Missing At Random)
  - Missingness a function of variables not in the model or latent variables in the model (such as the variable with missing)
  - NMAR modeling possible but difficult to know which model is best (Muthén, et al., 2011, Psych Methods) - sensitivity analysis

Ways to Check Missingness

- With respect to a key variable:
  - Compare the mean for a key variable using all observations on this variable versus using observations with no missing on a set of relevant variables (listwise)
    - The mean and variance for the outcome variable remove are not that different across the two samples
- With respect to predictors of missingness on a key variable
  - Do logistic or probit regression for a binary missing data indicator
    - Need for adding control variables?
- With respect to a model:
  - Compare key parameter estimates and SEs
    - Are differences substantively important?

Logistic Regression for Missing on Agg5

Table: Predicting from Covariates Including Black

	USEVARIABLES = tx agg1 black missing;
	CATEGORICAL = MISSING;
DEFINE:	IF (agg5 EQ _MISSING) THEN missing=1 ELSE missing=0;
ANALYSIS:	ESTIMATOR = ML; ! logistic regression
MODEL:	missing ON tx agg1 black;

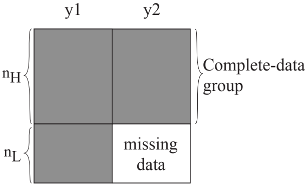
Table: Including Agg5 as a (Partly Latent) Predictor

	USEVARIABLES = tx agg1 agg5 black missing;
	CATEGORICAL = MISSING;
DEFINE:	IF (agg5 EQ _MISSING) THEN missing=1 ELSE missing=0;
ANALYSIS:	ESTIMATOR = ML; ! logistic regression
	INTEGRATION = MONTECARLO(500);
MODEL:	**missing ON agg5 tx agg1 black;**
	agg5 tx agg1 black; ! bringing the xs into the model

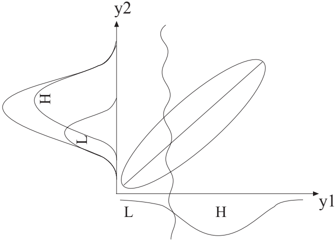
From Table 10.25: 16 missing data approaches for the aggression mediation example (indirect effects changed to STDY version).

Approach	remove ON agg5	agg5 ON tx	Indirect effect (STDY)
1. Listwise deletion (n = 250)	0.820 (.157)	-0.294 (.129)	[-0.237 -0.118 -0.010]
7. ML assuming MAR including subjects missing on agg1 and agg5 (n=441)	0.773 (.143)	-0.269 (.119)	[-0.090 -0.045 -0.006]
12. ML assuming MAR including subjects missing on agg1 and agg5 and adding black (n=441)	0.746 (.147)	-0.300 (.118)	[-0.093 -0.049 -0.011]

Missing Data Analysis Using ML Under MAR in a Bivariate Normal Case



(a)



(b)

ML estimates (using all available data):

$$\hat{\mu}_{y_2} = \hat{\mu}^*_{y_2} + \hat{\beta}^* \left(\hat{\mu}_{y_1} - \hat{\mu}^*_{y_1}\right),$$
$$\hat{\sigma}_{y_2,y_2} = \hat{\sigma}^*_{y_2,y_2} + \hat{\beta}^{*2} \left(\hat{\sigma}_{y_1,y_1} - \hat{\sigma}^*_{y_1,y_1}\right).$$

Asterisks denote complete-data (listwise) estimates

MAR for Mediation: Missing on Y as a Function of M

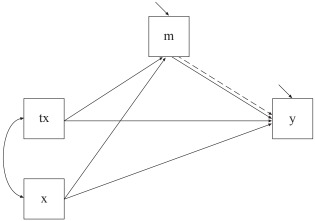


Figure: Data-Generating Model

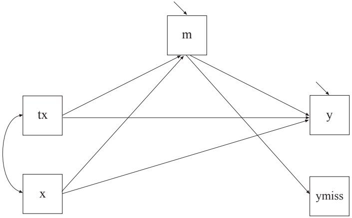


Figure: Full Analysis Model

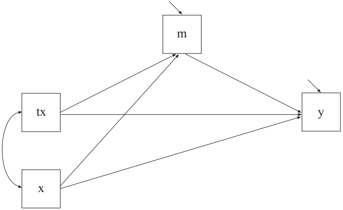
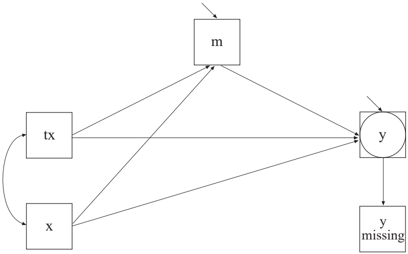


Figure: Simplified Analysis Model - Same as Full Model When MAR Holds

Figure: Missing data as a function of the latent outcome which corresponds to an observed variable with missing data



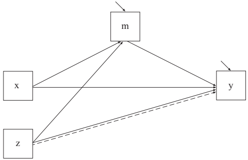


Figure: Missing data predictor *z* influences missingness and plays a substantive role

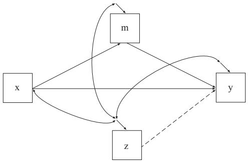
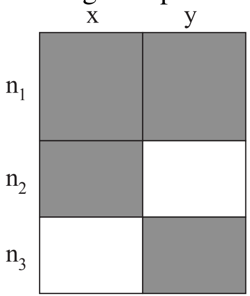


Figure: Missing data predictor *z* included in the model as a missing data correlate: Automated by AUXILIARY = z (M)

Bringing Xs Into the Model By Mentioning Them: When Does it Make a Difference?

Missing data patterns (blank is missing)



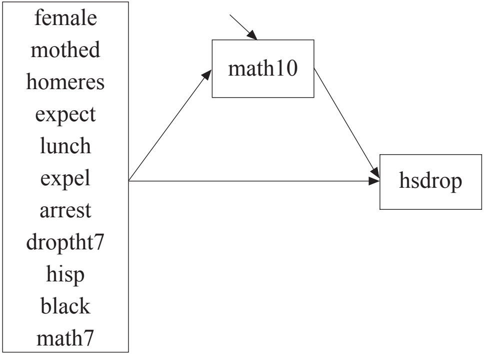
$$
\begin{aligned}
logL &= \sum_i log[y_i, x_i] \\
&= \sum_{i=1}^{n_1} log[y_i \mid x_i] \qquad (n_1) \\
&+ \sum_{i=1}^{n_1+n_2} log[x_i] \qquad (n_1+n_2) \\
&+ \sum_{i=n_1+n_2+1}^{n_1+n_2+n_3} log[y_i]. \quad (n_3)
\end{aligned}
$$

- The slope using the first term (n_1) doesn't change when adding the second term (n_2) - the sample size is only cosmetically bigger
- Adding the third term (n_3) changes the slope (a larger n is used)

The Danger of Bringing Xs Into the Model

- Too easy: Simply add the MODEL line x1-x10;
- Problem: Often a large amount of missing data on x's
  - **Too much reliance on the model relative to the data**
- Normality assumption for x's not always realistic or good enough
  - Binary x's can be treated as binary
  - Chapter 10 has a simulation study showing the benefit of using Bayes with binary covariates

Figure: Mediation model for a binary outcome of dropping out of high school (n=2898)



Missing On The Mediator And The Covariates Treating All Covariates As **Normal**: ML Versus Bayes

- ML requires integration over 10 dimensions (the mediator and 9 covariates have missing data and y is binary)
- ML needs 2,500 Monte Carlo integration points for sufficient precision
- ML needs bootstrap to represent the non-symmetric confidence interval for the indirect effect
- ML takes 6 hours with 1,000 bootstraps

- Bayes takes less than a minute
- Bayes doesn't need bootstrap because the non-symmetric CI is obtained as percentiles from the posterior distribution
- Bayes posterior based on 20,000 draws as compared to 1,000 bootstraps for ML

Treating Binary Covariates As Binary: ML Versus Bayes

6 covariates are binary and several represent rare events.

- ML requires 10 + 15 = 35 dimensions of integration: intractable (15 = 6*5/2 for 6 binary covariates where each pair needs a factor to represent their covariance)
- Bayes takes 3 minutes for 20,000 draws (multivariate normal model underlying the binary covariates captures the covariances - like for WLSMV probit)

Input for High School Dropout Mediation Analysis Treating Binary Covariates as Binary using Bayes

	MISSING = ALL(9999);
	CATEGORICAL = hsdrop female expel arrest droptht7 hisp black;
ANALYSIS:	ESTIMATOR = BAYES;
	PROCESSORS = 2;
	BITERATIONS = (20000);
	PREDICTOR = OBSERVED;
MODEL:	hsdrop ON math10 female-math7;
	math10 ON female-math7;
	female-math7 WITH female-math7;
MODEL INDIRECT:	hsdrop IND math10 math7(61.01 50.88);
OUTPUT:	PATTERNS TECH1 TECH8 CINTERVAL;
PLOT:	TYPE = PLOT3;

The Mplus User's Guide has Gotten a Companion