Mediation Models

BIOS 6611

CU Anschutz

Week 12

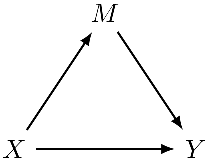
1. Mediation Analysis

2. A Few Extra Notes

Mediation Analysis

Mediators

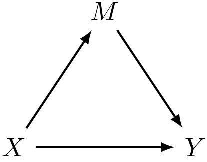
A **mediator** (M) is an intermediate variable on the causal path between an independent variable (e.g., treatment or exposure) and a dependent variable (i.e., outcome):



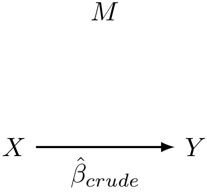
This relationship can be broken down into **indirect** and **direct effects**:

- Indirect Effect: effect of X on Y that works through M
- Direct Effect: effect of X on Y that does not work through M (the "remaining" effect after adjusting for the mediator)

Fundamental Models of Mediation Analysis



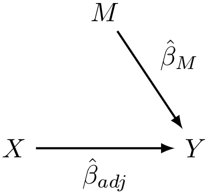
Crude Model



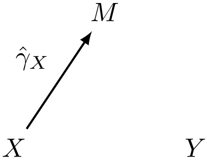
Our corresponding fitted regression models are

Crude Model:
$\hat{Y} = \hat{\beta}_{01} + \hat{\beta}_{crude} X$

Adjusted Model



Covariate Model



Adjusted Model:
$\hat{Y} = \hat{\beta}_{02} + \hat{\beta}_{adj} X + \hat{\beta}_M M$

Covariate Model:
$\hat{M} = \hat{\gamma}_0 + \hat{\gamma}_X X$

Inference for Mediation

The three fundamental models for mediation analysis are the same that we saw for confounding. However, for mediation, we can evaluate the significance of the indirect effect.

Like before, $\hat{\beta}_{crude} - \hat{\beta}_{adj} = \hat{\gamma}_X \times \hat{\beta}_M$. Further, the standard error can be derived using the delta method as

$$SE(\hat{\beta}_{crude} - \hat{\beta}_{adj}) = SE(\hat{\gamma}_X \times \hat{\beta}_M) = \sqrt{\hat{\beta}_M^2 Var(\hat{\gamma}_X) + \hat{\gamma}_X^2 Var(\hat{\beta}_M)}$$

Often we summarize the **proportion mediated** by M as

$$\frac{\text{indirect effect}}{\text{total effect}} = \frac{\hat{\beta}_{crude} - \hat{\beta}_{adj}}{\hat{\beta}_{crude}} = \frac{\hat{\gamma}_X \times \hat{\beta}_M}{\hat{\beta}_{crude}}$$

Multiplying by 100 results in the percent mediated.

Inference for Mediation

In addition to the proportion mediated summary statistic, we can also calculate a 95% confidence interval and p-value based on the indirect effect.

For the 95% confidence interval around the indirect effect:

$$(\hat{\gamma}_X \times \hat{\beta}_M) \pm Z_{\alpha/2} SE(\hat{\gamma}_X \times \hat{\beta}_M)$$

On the proportion/percent mediated scale, we divide the lower and upper estimate by the total effect.

For a Z-statistic:

$$Z = \frac{\text{indirect effect}}{SE(\hat{\gamma}_X \times \hat{\beta}_M)},$$

where a p-value can be calculated by referencing the standard normal distribution.

Mediation Example

A study was performed to examine if a weight loss drug could cause a decrease in systolic blood pressure in adolescents with severe obesity that was already demonstrated in adults.

It was hypothesized that changes in BMI would account for the decrease in blood pressure observed in the subjects on the study drug.

21 adolescents were randomized in total: 9 to the treatment group and 12 to the control group.

The data set includes group (`trtgrp`: $1 =$ drug, $0 =$ placebo), change in BMI (`cbmi`), and change in systolic blood pressure (`csbp`).

Our DAG in this case would be:

Model 1 - Crude Model

```
dat <- read.csv('pilot_top.csv')
crude_model <- glm( csbp ~ trtgrp, data=dat )
summary(crude_model)$coefficients
```

```
##              Estimate Std. Error    t value   Pr(>|t|)
## (Intercept) -1.916667   2.202277 -0.8703113 0.39499327
## trtgrp      -8.972222   3.364034 -2.6671023 0.01523101
```

Interpretation: There is an 8.97 mmHg greater reduction in systolic blood pressure in the treatment group compared to the control group. This represents a significant difference between the two groups (p=0.015).

Is confounding likely in this study? No, since the groups were randomized we'd expect factors be balanced across groups.

Model 2 - Adjusted Model

```r
adjusted_model <- glm( csbp ~ trtgrp + cbmi, data=dat )
summary(adjusted_model)$coefficients
```

```
##               Estimate Std. Error   t value  Pr(>|t|)
## (Intercept) -1.4457035  2.2879823 -0.6318683 0.5354165
## trtgrp      -4.5451684  6.2403990 -0.7283458 0.4757715
## cbmi         0.8073654  0.9555404  0.8449307 0.4092403
```

Interpretation: After adjusting for change in BMI, there is a 4.55 mmHg greater reduction in systolic blood pressure in the treatment group compared to the control group. This difference is not significantly different than zero (p=0.476).

Indirect Effect: Based on our crude and adjust model we can estimate our indirect effect: $\hat{\beta}_{\text{crude}} - \hat{\beta}_{\text{adj}} = -8.972 - (-4.5452) = -4.4268$.

Model 3 - Covariate Model

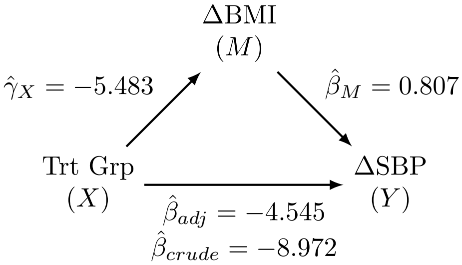
```
covariate_model <- glm( cbmi ~ trtgrp, data=dat )
summary(covariate_model)$coefficients
```

```
##               Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) -0.5833333  0.5327711 -1.094904 2.872447e-01
## trtgrp      -5.4833333  0.8138213 -6.737761 1.942590e-06
```

Indirect Effect: We can also verify that the indirect effect calculated from the adjusted and covariate models equals what we found on our previous slide: $\hat{\gamma}_X \times \hat{\beta}_M = (-5.4833)x(0.8074) = -4.4268$.

In our mediation context, the indirect effect is the extent to which the outcome (ΔSBP) changes when the PEV (treatment group) is held fixed and the mediator variable (ΔBMI) changes by the amount it would have changed had the PEV increased by one unit.

Proportion Mediated



$$\frac{\text{indirect effect}}{\text{total effect}} = \frac{\hat{\beta}_{crude} - \hat{\beta}_{adj}}{\hat{\beta}_{crude}} = \frac{\hat{\gamma}_X \times \hat{\beta}_M}{\hat{\beta}_{crude}} = \frac{-4.4268}{-8.972} = 0.4934$$

Interpretation: 49.3% of the effect of the study drug on changes in systolic blood pressure is mediated through (can be explained by) changes in BMI.

Proportion Mediated: Test Statistic and p-value

We first need to calculate the standard error:

$$
\begin{aligned}
SE\left(\hat{\gamma}_X \times \hat{\beta}_M\right) &= \sqrt{\hat{\gamma}_X^2 \left(SE\left(\hat{\beta}_M\right)\right)^2 + \hat{\beta}_M^2 \left(SE\left(\hat{\gamma}_X\right)\right)^2} \\
&= \sqrt{(-5.48)^2 (0.9555)^2 + (0.807)^2 (0.8138)^2} \\
&= 5.277164
\end{aligned}
$$

The test statistic is $Z = \frac{\text{indirect effect}}{SE\left(\hat{\gamma}_X \times \hat{\beta}_M\right)} = \frac{-4.4268}{5.277164} = -0.839$.

The p-value is $p =$`2*pnorm(-0.839)`$=$0.4015.

Proportion Mediated: CI and Conclusion

The 95% CI is $-4.4268 \pm 1.96 \times 5.277164 = (-9.704, 0.850)$.

In the context of the proportion mediated, the 95% CI is $\left(\frac{-9.704}{-8.972}, \frac{0.850}{-8.972}\right) = (-0.095, 1.082)$.

This represents an *inconsistent mediation model* (more on next slide).

Interpretation: Change in BMI is not a significant mediator of the relationship between treatment and change in systolic blood pressure (p=0.401), even though it explains 49.3% (95% CI: -9.5% to 108.2%) of the effect of treatment in our sample.

Inconsistent Mediation

What we discussed is one of the most basic formulations of the mediation framework for hypothesis testing, sometimes called *Sobel's test*. There are more complex approaches that may be more appropriate for different situations.

For example, it is possible with our simpler approach to arrive at a proportion mediated that is negative or greater than 100%. This can occur with *inconsistent (mediation) models* where at least one mediated effect has a different sign than the other mediated or direct effects in the model.

The proportion mediated may also be unstable for sample sizes under 500.

Freedman LS. Confidence intervals and statistical power of the "validation" ratio for surrogate or intermediate endpoints. *J Statist Plan Inference*. 2001;96:143–53.

MacKinnon DP, Warsi G, Dwyer JH. A simulation study of mediated effect measures. *Multivariate Behav Res*. 1995;30:41–62.

A Few Extra Notes

Standard Error of the Indirect Effect

For a random variable X, the **delta method** can be used to derive the approximate variance of a function of a random variable, $f(X)$, as

$$Var[f(X)] = [f'(X)]^2 Var(X)$$

How is this variance derived? Recall from calculus, the first order Taylor series approximation of a function of X that is expanded about μ, the mean of X, is

$$g(X) = g(\mu) + (X - \mu)g'(\mu)$$

The variance of $g(X)$, the function of X, can then be estimated as:

$$Var[g(x)] = Var[g(u)] + [g'(\mu)]^2 Var[X - \mu] \approx [g'(\mu)]^2 \sigma^2$$

The variance of a function of k random variables $f(X)$ is approximated by:

$$Var[f(x)] = \sum_{i=1}^{k} f_i'(X)^2 Var(X_i) + 2 \sum_{i>j} f_i'(X) f_j'(X) Cov(X_i, X_j)$$

Sobel used this approach to derive the standard error.

Bootstrapping Instead of Normal Approximations

One alternative that avoids the issues of Sobel's approximation for the variance is to use bootstrap sampling to evaluate the variability around our estimator (e.g., the indirect effect or the proportion mediated).

This avoids any distributional assumption and may produce more accurate estimates of our confidence interval.

Bootstrap sampling, however, will not correct instability of the estimated indirect effect for small samples (e.g., $<$500). Therefore we still may end up with inconsistent mediation models and wide intervals with smaller samples.