

4/22/2010

36-402/608 ADA-II
Breakout #24: Mediation 2

H. Seltman

Job Search Intervention Study (JOBS II) is a randomized field experiment that investigates the efficacy of a job training intervention on unemployed workers. The program is designed to not only increase reemployment among the unemployed but also enhance the mental health of the job seekers. In the JOBS II field experiment, 1,801 unemployed workers received a pre-screening questionnaire and were then randomly assigned to treatment and control groups. Those in the treatment group participated in job-skills workshops. In the workshops, respondents learned job-search skills and coping strategies for dealing with setbacks in the job-search process. Those in the control condition received a booklet describing job-search tips. In follow-up interviews, the two key outcome variables were measured; a continuous measure of depressive symptoms based on the Hopkins Symptom Checklist which gives a value between 0 and 5 with 2 decimal places, and a binary variable, representing whether the respondent had become employed.

```
library("mediation")
data("jobs")
summary(jobs)
```

treat	econ_hard	depress1	sex	age
Min. :0.0000	Min. :1.000	Min. :1.00	Min. :0.0000	Min. :17.49
1st Qu.:0.0000	1st Qu.:2.330	1st Qu.:1.36	1st Qu.:0.0000	1st Qu.:29.30
Median :1.0000	Median :3.000	Median :1.83	Median :1.0000	Median :36.64
Mean :0.6674	Mean :3.024	Mean :1.87	Mean :0.5362	Mean :37.57
3rd Qu.:1.0000	3rd Qu.:3.670	3rd Qu.:2.36	3rd Qu.:1.0000	3rd Qu.:44.62
Max. :1.0000	Max. :5.000	Max. :3.00	Max. :1.0000	Max. :72.48

	occp	marital	nonwhite	educ
professionals	:175	nev marr:279	white0 :747	lt-hs : 50
manegerial	:168	married:408	non.white1:152	highsc:272
clerical/kindred	:217	separtd: 30		somcol:319
sales workers	: 65	divrcd :163		bach :146
craftsmen/foremen/kindred	: 97	widowed: 19		gradwk:112
operatives/kindred wrks	: 93			
laborers/service wrks	: 84			

income	job_seek	depress2	work1	job_dich
lt15k :164	Min. :1.000	Min. :1.000	psyump:606	Min. :0.0000
15t24k:206	1st Qu.:3.667	1st Qu.:1.273	psyemp:293	1st Qu.:0.0000
25t39k:218	Median :4.167	Median :1.600		Median :1.0000
40t49k:110	Mean :4.043	Mean :1.741		Mean :0.6174
50k+ :201	3rd Qu.:4.667	3rd Qu.:2.091		3rd Qu.:1.0000
	Max. :5.000	Max. :4.909		Max. :1.0000

control	job_disc
Min. :0.0000	Low-Med : 24
1st Qu.:0.0000	Medium :138
Median :0.0000	Med-High:438
Mean :0.3326	High :299
3rd Qu.:1.0000	
Max. :1.0000	

Consider the model to test whether the effects of the workshops (treat) on depression at the time of follow-up (depress2) is mediated through increasing job seeking actions (job_seek, considered to be on a quantitative scale), correcting for pretreatment level of depression (depress1), education, income, race, marital status, age, sex (a female indicator), previous occupation (occp), and the level of economic hardship (econ_hard).

Question 1: Draw the DAG.

```

model.m <- lm(job_seek ~ treat + depress1 + econ_hard + sex +
              age + occp + marital + nonwhite + educ + income,
              data = jobs)

summary(model.m)
# Coefficients:              Estimate Std. Error t value Pr(>|t|)
# (Intercept)              3.8806256  0.1947174  19.930 < 2e-16 ***
# treat                    0.0774238  0.0492939   1.571 0.116624
# depress1                 -0.2540256  0.0440638  -5.765 1.13e-08 ***
# econ_hard                0.1036040  0.0265612   3.901 0.000103 ***
# sex                     -0.0053180  0.0542355  -0.098 0.921913
# age                      0.0005308  0.0026550   0.200 0.841575
# occpmanegerial          0.0056477  0.0766773   0.074 0.941302
# occpclerical/kindred   -0.1132352  0.0777967  -1.456 0.145883
# occpsales workers      -0.0137738  0.1012484  -0.136 0.891821
# occpcraftsmen/foremen/kindred -0.2015647  0.0965720  -2.087 0.037160 *
# occpoperatives/kindred wrks -0.2959024  0.0999259  -2.961 0.003147 **
# occplaborers/service wrks -0.3565544  0.1019639  -3.497 0.000494 ***
# maritalmarried         0.0381422  0.0667623   0.571 0.567934
# maritalsepartd         0.3641314  0.1365291   2.667 0.007793 **
# maritaldivrcd          0.2041101  0.0770659   2.649 0.008230 **
# maritalwidowed        -0.3301324  0.1761481  -1.874 0.061240 .
# nonwhitenon.white1     0.0615794  0.0651346   0.945 0.344707
# educhighsc             0.1813264  0.1088818   1.665 0.096201 .
# educsomcol             0.1638371  0.1097146   1.493 0.135719
# educbach               0.2563072  0.1220038   2.101 0.035943 *
# educgradwk            0.2013935  0.1293041   1.558 0.119709
# income15t24k           0.1583888  0.0753071   2.103 0.035730 *
# income25t39k           0.0898314  0.0776033   1.158 0.247355
# income40t49k           0.1999402  0.0941763   2.123 0.034031 *
# income50k+             0.1631108  0.0888391   1.836 0.066694 .

```

Question 2: What does this model tell us? Which causal step is tested here?

```

model.y <- lm(depress2 ~ treat + job_seek + depress1 + econ_hard +
              sex + age + occp + marital + nonwhite +
              educ + income, data = jobs)

summary(model.y)
# Coefficients:                Estimate Std. Error t value Pr(>|t|)
# (Intercept)                1.4527281  0.1940639   7.486 1.74e-13 ***
# treat                       -0.0367886  0.0407940  -0.902  0.36740
# job_seek                    -0.1773802  0.0279535  -6.346 3.56e-10 ***
# depress1                     0.4098612  0.0371003  11.047 < 2e-16 ***
# econ_hard                    0.0679692  0.0221404   3.070  0.00221 **
# sex                          0.0621566  0.0448206   1.387  0.16586
# age                          0.0007858  0.0021942   0.358  0.72034
# occpmanegerial              0.0664879  0.0633665   1.049  0.29435
# occpclerical/kindred        0.0503458  0.0643692   0.782  0.43434
# occpsales workers           -0.0348333  0.0836727  -0.416  0.67729
# occpcraftsmen/foremen/kindred -0.0290567  0.0800059  -0.363  0.71656
# occpoperatives/kindred wrks  0.1635053  0.0829922   1.970  0.04914 *
# occplaborers/service wrks   -0.0215721  0.0848505  -0.254  0.79937
# maritalmarried              -0.0072627  0.0551828  -0.132  0.89532
# maritalsepartd              0.2019970  0.1132861   1.783  0.07492 .
# maritaldivrcd               -0.0453020  0.0639424  -0.708  0.47884
# maritalwidowed              0.0923133  0.1458613   0.633  0.52697
# nonwhitenon.white1         -0.1081444  0.0538549  -2.008  0.04494 *
# educhighsc                  -0.0023664  0.0901228  -0.026  0.97906
# educsomcol                   0.0226457  0.0907839   0.249  0.80307
# educbach                     0.0148269  0.1010784   0.147  0.88341
# educgradwk                   0.1782504  0.1070053   1.666  0.09611 .
# income15t24k                -0.0486597  0.0623912  -0.780  0.43565
# income25t39k                 -0.0208905  0.0641806  -0.325  0.74488
# income40t49k                 -0.0528838  0.0780279  -0.678  0.49811
# income50k+                   -0.1179727  0.0735582  -1.604  0.10912

```

Question 3: What does this model tell us? Which causal step is tested here?

The Sobel test defines a as the coefficient of T in the regression of M on T and X, and b as the coefficient of M in the regression of Y on M, T, and X. In the mediation literature the notation for the *sampling* variance of these coefficients (square of the standard errors of the coefficients) is s_a^2 and s_b^2 . The Sobel test is a Z-test of $H_0 : ab = 0$ using $SE_{ab} = \sqrt{a^2 s_b^2 + b^2 s_a^2 + s_a^2 s_b^2}$.

Question 4: Find the values needed for this formula in the above results. If you have access to the web, calculate the Sobel test with <http://people.ku.edu/~preacher/sobel/sobel.htm> (or use a calculator). Note that the sampling distribution of a product is often not normal, so this test may be unreliable.

Use the notation T for the treatment of interest, M for the mediator, X for pre-treatment covariates, and Y for the outcome. The actual mediation analysis is performed by the `mediate()` function using a model of M on T and X (called `model.m`), and a model of Y on M, T and X (called `model.y`). Models other than `lm()`, such as `glm()`, are allowed. Two methods are provided, but they tend to give similar results.

Important technical detail: “These two model objects, `model.m` and `model.y`, become the arguments for the `mediate()` function. The analyst must take some care with missing values before estimating the models above. While model functions in R handle missing values in the data using the usual listwise deletion procedures, the functions in mediation assume that missing values have been removed from the data before the estimation of these two models. Thus the data for the two models must have identical observations sorted in the same order with all missing values removed.”

```
out.1 <- mediate(model.m, model.y, sims = 1000, boot = TRUE,
                treat = "treat", mediator = "job_seek")
summary(out.1)
# Causal Mediation Analysis
# Confidence Intervals Based on Nonparametric Bootstrap
# Mediation Effect:  -0.01371 95% CI  -0.033558  0.002373
# Direct Effect:    -0.03779 95% CI  -0.1164  0.0374
# Total Effect:    -0.0515 95% CI  -0.13155  0.02468
# Proportion of Total Effect via Mediation:  0.2217 95% CI  -1.944  2.643
```

The “Mediation Effect” is the product ab , i.e., the (estimated) mediated effect of a one unit change in T on Y. The output labeled “Direct Effect” is the direct effect of T on Y, which would be zero in the case of complete mediation.

```
out.2 <- mediate(model.m, model.y, sims = 1000, treat = "treat",
                mediator = "job_seek")
summary(out.2)
# Causal Mediation Analysis
# Quasi-Bayesian Confidence Intervals
# Mediation Effect:  -0.01366 95% CI  -0.031616  0.002241
# Direct Effect:    -0.03821 95% CI  -0.11495  0.04142
# Total Effect:    -0.05187 95% CI  -0.13195  0.03282
# Proportion of Total Effect via Mediation:  0.2133 95% CI  -2.651  2.061
```

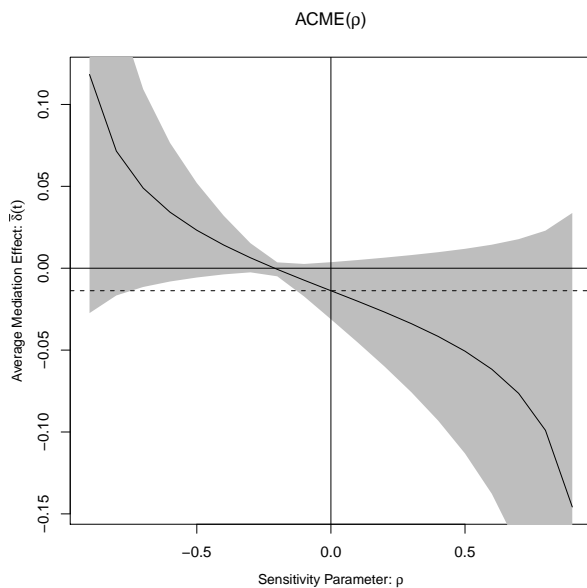
Question 5: Do we have evidence that job seeking actions mediate the effect of the workshops on depression at the time of follow-up?

A key assumption of causal mediation, even in the presence of randomized treatment assignment is that there is no unmeasured “Z” that affects (“causes”) both M and Y. It is often hoped that measuring sufficient pre-treatment covariates (X’s) will preclude the presence of any important unmeasured Z’s. The assumption can be expressed in the form $\rho = 0$ where ρ is the correlation of ϵ_1 and ϵ_2 in the equations

$$\begin{aligned} M_i &= T_i\beta + X_i\gamma + \epsilon_{i1} \\ Y_i &= T_i\beta + X_i\gamma + M_i\theta + \epsilon_{i2} \end{aligned}$$

The assumption is untestable, but we can perform “sensitivity analysis”.

```
plot(medsens(out.1))
```



Question 6: ACME means Average Causal Mediation Effect. Interpret the plot. Give plausible examples of a unmeasured confounders that induce positive and negative error correlations.

```

model.yw <- glm(work1 ~ job_seek + treat + depress1 + econ_hard +
               sex + age + occp + marital + nonwhite + educ + income,
               data = jobs, family=binomial)
summary(model.yw)
# Coefficients:                Estimate Std. Error z value Pr(>|z|)
# job_seek                    0.228010   0.109752   2.077   0.0378 *
# treat                       0.279319   0.160911   1.736   0.0826
# ...
out.w <- mediate(model.m, model.yw, sims = 1000, treat = "treat",
                 mediator = "job_seek")
summary(out.w)
# Mediation Effect:  0.003666 95% CI -0.0009565  0.0108308
# Direct Effect:    0.05476 95% CI -0.006568  0.119461
# Total Effect:     0.05843 95% CI -0.003512  0.124070
# Proportion of Total Effect via Mediation:  0.0576 95% CI -0.2573  0.3555

```

Question 7: How does this analysis differ from above? Important note: mediate() is valid in this case, but the Sobel test based on the usual SE formula is not.