Mediation analysis in Stata: a short look at med4way

Rino Bellocco

Department of Statistics and Quantitative Methods, University of Milano-Bicocca
Departments of Medical Epidemiology and Biostatistics, Karolinska Institutet

rino.bellocco@ki.se

February 18, 2021

Outline

1. med4way

Outline

Outline

Outline

Summary

Acknowledgment

- **Discacciati, A.**, Bellavia, A., Lee, J.J., Mazumdar, M., Valeri, L. Med4way: a Stata command to investigate mediating and interactive mechanisms using the four-way effect decomposition. *International Journal of Epidemiology*. 2019 Feb;48(1):15-20.
- A Stata command for the 4-way decomposition using parametric regression models
- https://github.com/anddis/med4way
- net install med4way, from("https://raw.githubusercontent.com/anddis/med4way/master/") replace

In nutshell

- help med4way
- current version: **v2.3.1** - **25jul2019**
- uses parametric regression models to estimate the components of the 4-way decomposition of the total effect of an exposure on a outcome in the presence of a mediator with which the exposure may interact. This decomposition breaks down the total effect of the exposure on the outcome into components due to mediation alone, to interaction alone, to both mediation and interaction, and to neither mediation nor interaction
- confidence intervals for the derived effects
- Improves computational speed by using the delta method for calculating standard errors components using the delta method (default) or the bootstrap
- allows continuous, binary, count or **survival outcomes**, and continuous or binary mediators

Summary

Models

Two regression models are fitted:

- model for the outcome (as a function of the exposure, the mediator, their interaction and confounders)
  - linear
  - logistic, log-binomial, Poisson, negative binomial
  - accelerated failure time (exponential, Weibull) and Cox
- model for the mediator (as a function of the exposure and confounders
  - linear
  - logistic

- the causal effects are automatically computed by the command as a function of the regression parameters estimated from the above specified models.

Computation

- two functions written in Mata language
- to calculate the components of the four-way decomposition, the numerical derivatives and the matrix multiplications necessary for the delta method.
- med4way follows standard Stata language syntax
- The main argument of the command is a list of variables, which has to follow a specific order: outcome, exposure, mediator and, if any, confounder variables

Important:

- In the case of a survival outcome, the outcome variable must be omitted.
- data must be read as survival time, using Stata's stset command,
- med4way is fully integrated with Statas way of handling survival data
- The variable for the interaction between the exposure and the mediator is automatically generated and added to the model for the outcome.

command

med4way [yvar] avar mvar [cvars], **a0**(#) **a1**(#) m(#) yreg() mreg()

- **a0()** specifies the referent level of the exposure;
- **a1()** specifies the actual level of the exposure;
- **m()** specifies the level of the mediator at which the fourway decomposition is computed
- **yreg()** specifies the form of the regression model for the outcome
- **mreg()** specifies the form of the regression model for the mediator

other options

- **c()** fixes the values of the confounders
- **bootstrap**
- **fulloutput**
- **casecontrol**

postestimation

After running med4way, a suite of user-friendly official Stata
post-estimation commands is readily available

- lincom, nlcom
- test, testnl
- estat bootstrap
- estat vce

Summary

Social Inequalities in Cancer Survival

The aim is to understand *if* and *to which extent* the effect of *SEP* (socio-economic position) inequalities on cancer mortality is mediated through stage at diagnosis (advanced versus non advanced)

- Data for Surveillance Epidemiology End Results (SEER) linked to American Community Survey (ACS) data for patients diagnosed in 1992-2005 and followed up to 2010
- In the present analysis, interaction between the SEP measure and stage at diagnosis was allowed
- Potential confounders: gender, age at diagnosis, year at diagnosis and state of residence
- Since the outcome is failure time a censoring variable was define taking value 1 if the individual is censored or value 0 if the event is observed
- The survival outcome is studied using an accelerated failure time model assuming a Weibull distribution

Social Inequalities in Cancer Survival

```
. stset stset srv_time_mon_pa, failure(censor)

. med4way new_medianincome stage_dich_noinsitu sex rac_reca ///
  age_c date_c ///
  state_2 state_3 state_4 state_5 state_6 state_7 state_8, ///
  a0(25000) a1(75000) m(1) yreg(aft, we) mreg(logistic)

. med4way new_medianincome stage_dich_noinsitu sex rac_reca ///
  age_c date_c ///
  state_2 state_3 state_4 state_5 state_6 state_7 state_8, ///
  a0(25000) a1(75000) m(0) yreg(aft, we) mreg(logistic)
```

Social Inequalities in Cancer Survival

```
Summary

  Outcome    (yvar):  srv_time_mon_pa
  Exposure   (avar):  new_medianincome
  Mediator   (mvar):  stage_dich_noinsitu
  Covariates (cvars): sex rac_reca age_c date_c state_2
  state_3 state_4 state_5 state_6 state_7 state_8

  Model for the outcome  (yreg): aft, weibull
  Model for the mediator (mreg): logistic

  Referent exposure level (a0):            25000
  Actual exposure level   (a1):            75000
  Mediator level for the decomposition (m): 1
```

Social Inequalities in Cancer Survival: Outcome Model

```
Weibull AFT regression

No. of subjects =      93,797      Number of obs    =      93,797
No. of failures =      62,738
Time at risk    =    6716569
                                   LR chi2(14)      =    35775.51
Log likelihood  =  -136942.97      Prob > chi2      =     0.0000

---------------------------------------------------------------------------------
                      _t |   Coef. Std. Err.     z  P>|z|   [95% Conf. Interval]
-------------------------+-------------------------------------------------------
          new_medianincome |  0.000006 0.000001    9.08 0.000   0.000005    0.000007
      stage_dich_noinsitu |-1.403908 0.042142  -33.31 0.000  -1.486505   -1.321310
_new_medianinXstage_dich_~000 |-0.000008 0.000001   -8.63 0.000  -0.000009   -0.000006
                     sex |  0.176502 0.009440   18.70 0.000   0.157999    0.195005
                rac_reca |-0.213340 0.015855  -13.46 0.000  -0.244415   -0.182265
                   age_c |-0.048775 0.000425 -114.65 0.000  -0.049609   -0.047941
                   _cons |  4.647863 0.037425  124.19 0.000   4.574513    4.721214

-------------------------+-------------------------------------------------------

                   /ln_p | -0.154948 0.003305  -46.89 0.000  -0.161425   -0.148471
-------------------------+-------------------------------------------------------
                       p |  0.856460 0.002830                0.850930    0.862025
                     1/p |  1.167597 0.003859                1.160059    1.175184
---------------------------------------------------------------------------------
```

Social Inequalities in Cancer Survival: Mediator Model

```
 Model for the mediator

Iteration 0:   log likelihood = -55851.309
Iteration 1:   log likelihood = -55678.911
Iteration 2:   log likelihood = -55678.204
Iteration 3:   log likelihood = -55678.204

Logistic regression                          Number of obs    =     100,000
                                             LR chi2(12)      =      346.21
                                             Prob > chi2      =      0.0000
Log likelihood = -55678.204                  Pseudo R2        =      0.0031

----------------------------------------------------------------------------------
stage_dich_noinsitu |     Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
--------------------+-------------------------------------------------------------
   new_medianincome |  -0.000003   0.000001    -3.17   0.002    -0.000005   -0.000001
                sex |   0.000180   0.014822     0.01   0.990    -0.028872    0.029231
           rac_reca |   0.331178   0.023816    13.91   0.000     0.284501    0.377856
              age_c |   0.003913   0.000572     6.84   0.000     0.002792    0.005035
              _cons |  -1.041371   0.053634   -19.42   0.000    -1.146492   -0.936249
----------------------------------------------------------------------------------
```

m=1

4-way decomposition: delta method

```
-------------------------------------------------------------------------------
             |     Coef.   Std. Err.      z     P>|z|     [95% Conf. Interval]
-------------+-----------------------------------------------------------------
      tereri |   0.356936   0.044769    7.97    0.000      0.269191    0.444682
   ereri_cde |  -0.020345   0.010062   -2.02    0.043     -0.040066   -0.000624
ereri_intref |   0.338726   0.039431    8.59    0.000      0.261443    0.416009
ereri_intmed |   0.012057   0.004086    2.95    0.003      0.004049    0.020065
   ereri_pie |   0.026498   0.008436    3.14    0.002      0.009965    0.043031
-------------------------------------------------------------------------------
```

m=0

```
 4-way decomposition: delta method

> 4-way decomposition: delta method

------------------------------------------------------------------------------
            |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
------------+-----------------------------------------------------------------
     tereri |   0.356936   0.044769    7.97   0.000     0.269191    0.444682
  ereri_cde |   0.435968   0.055314    7.88   0.000     0.327554    0.544381
ereri_intref| -0.117587   0.013886   -8.47   0.000    -0.144802   -0.090371
ereri_intmed|  0.012057   0.004086    2.95   0.003     0.004049    0.020065
  ereri_pie |   0.026498   0.008436    3.14   0.002     0.009965    0.043031
------------------------------------------------------------------------------
```

Intepretation - Total Effect

- The survival analysis yielded a positive effect of income on survival, a negative effect of stage
- A negative, significant interaction between tumor stage at diagnosis and county household median income was detected
- The logistic regression analysis showed a negative association between the SEP measure and stage at diagnosis
- a total effect of 1.35 indicating that the mean survival time of individuals living in counties with median income of 75,000 was higher than that of individuals living in counties with median income of 25,000 in the mean survival ratio scale (total excess relative risk = TE-1 = 0.35)

Intepetation - direct and indirect effect

- The controlled direct effect, controlling the mediator at level m=1 reveals that, had we intervened setting stage at diagnosis to be advanced for all individuals, mean survival time of individuals living in counties with median income of 75,000 would be lower than that of individuals living in counties with median income of 25,000, with a relative excess risk due to controlled direct effect estimated as -0.02

- An estimate of 0.34 for relative excess risk to interaction indicates that income disparities were higher among individuals with advanced stage at diagnosis.

- The direct effect and interaction only effect accounted for most of the observed total effect (90%)

Intepretation - direct and indirect effect

- The relative excess risk due to mediated interaction and pure indirect effect were estimated as 0.012 and 0.026 respectively, indicating that a mediating mechanism through stage explained about 10% of the disparity
- In explaining the mechanisms that lead to income disparities in survival outcomes stage at diagnosis appears to be involved via both interactive and mediating mechanisms.
- In explaining the mechanisms that lead to income disparities in survival outcomes stage at diagnosis appears to be involved via both interactive and mediating mechanisms. The former role seems more important that the latter

Summary

- The first Stata command for the decomposition of the total effect in mediating and interactive effects allowing for continuous, binary, count and survival outcome and for a continuous or binary mediator
- The command accommodates both cohort and case-control designs
- Methods for missing data and measurement error that might induce bias in the analyses are be incorporated
- developed a decomposition of the total effect in the presence of multiple mediators.