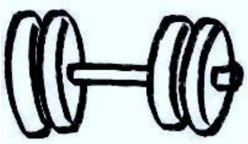
Generating survival data for fitting marginal structural Cox models using Stata

Presenter: **Ehsan Karim**
Department of Statistics
University of British Columbia
ehsan@stat.ubc.ca

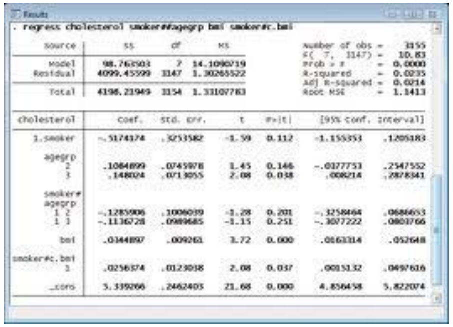
July 27, 2012
2012 Stata Conference in San Diego, California

Outline

- Idea of MSM



- Various weights



```
. regress cholesterol smoker#agegrp bmi smoker#c.bmi

      Source |       SS       df       MS              Number of obs =    3155
-------------+------------------------------           F(  7,  3147) =   10.83
       Model |  98.763503     7  14.1090719            Prob > F      =  0.0000
    Residual | 4099.45599  3147  1.30265522            R-squared     =  0.0235
-------------+------------------------------           Adj R-squared =  0.0214
       Total | 4198.21949  3154  1.33107783            Root MSE      =  1.1413

------------------------------------------------------------------------------
 cholesterol |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    1.smoker |  -.5174174   .3253582    -1.59   0.112    -1.155353    .1205183
             |
      agegrp |
          2  |   .1084899   .0745978     1.45   0.146    -.0377753    .2547552
          3  |    .148024   .0713055     2.08   0.038      .008214    .2878341
             |
     smoker# |
      agegrp |
        1 2  |  -.1285906   .1006039    -1.28   0.201    -.3258464    .0686653
        1 3  |  -.1136728   .0989685    -1.15   0.251    -.3077222    .0803766
             |
         bmi |   .0344897    .009261     3.72   0.000     .0163314     .052648
             |
smoker#c.bmi |
          1  |   .0256374   .0123038     2.08   0.037     .0015132    .0497616
             |
       _cons |   5.339266   .2462403    21.68   0.000     4.856458    5.822074
------------------------------------------------------------------------------
```

- Fitting MSM in Stata
  - using pooled logistic
  - using CoxPH (proposed)

- Simulation and data generation in Stata
- Stata vs. SAS/R

Idea of MSM

Y = outcome
A = treatment

Observed data stratified by confounder L:

	L = 1		L = 0	
	A = 1	**A = 0**	**A = 1**	**A = 0**
Y = 1	150	45	20	5
Y = 0	300	10	40	55
Total	**450**	**55**	**60**	**60**

Merged data:

	A = 1	A = 0
Y = 1	170	50
Y = 0	340	65
Total	**510**	**115**

Idea of MSM

- do http://stat.ubc.ca/~e.karim/research/pointmsm.do

- mata: data = tabc(150, 45, 20, 5, 300, 10, 40, 55, w = 0, s = 0, n = 0)
- mata: st_matrix("data",data)
- svmat double data, name(data)
- renvars data1-data5\ L A Y N w

- mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 0, s = 0, n = 0)

Idea of MSM

- **mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 0, s = 0, n = 0)**

```
                   1                2                3
    +-----------------------------------------------+
  1 |   -.1014492754     .7666666667             .65  |
    +-----------------------------------------------+
```

Risk difference   Risk Ratio   Odds Ratio

	L	A	Y	N
1	1	1	1	150
2	1	1	0	300
3	1	0	1	45
4	1	0	0	10
5	0	1	1	20
6	0	1	0	40
7	0	0	1	5
8	0	0	0	55

	A = 1	A = 0
Y = 1	170	50
Y = 0	340	65
Total	**510**	**115**

Idea of MSM

Ref: Robins et al. (2000)

- **mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 1, s = 0, n = 0)**

```
                  1                2                3
    +-------------------------------------------------+
  1 |   -.3437575758      .492302184      .238453276  |
    +-------------------------------------------------+
```

Risk difference Risk Ratio Odds Ratio

W = 1/P(A|L)

	L	A	Y	N	w	Nw
1	1	1	1	150	1.1222222	168.33333
2	1	1	0	300	1.1222222	336.66667
3	1	0	1	45	9.1818182	413.18182
4	1	0	0	10	9.1818182	91.818182
5	0	1	1	20	2	40
6	0	1	0	40	2	80
7	0	0	1	5	2	10
8	0	0	0	55	2	110

w	A = 1	A = 0
Y = 1	208	423
Y = 0	417	202
Total	**625**	**625**

Various weights

Ref: Hernán et al. (2002)
Xiao et al. (2010)

w = weighted?
s = stabilized?
n = normalized?

Unweighted: W = 1

- **mata: causal(..., w = 0, s = 0, n = 0)**

Simple weight: W = 1/P(A|L)

- **mata: causal(..., w = 1, s = 0, n = 0)**

Normalized weight: Wn = W/mean$_{_risk\ set}$(W)

- **mata: causal(..., w = 1, s = 0, n = 1)**

Stabilized weight: SW = P(A)/P(A|L)

- **mata: causal(..., w = 1, s = 1, n = 0)**

Normalized stabilized weight: SWn = SW/mean$_{_risk\ set}$ (SW)

- **mata: causal(..., w = 1, s = 1, n = 1)**

Various weights

Ref: Hernán et al. (2002)
Xiao et al. (2010)

```
mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 0, s = 0, n = 0)
                  1               2               3
    +-----------------------------------------------+
  1 |  -.1014492754    .7666666667             .65  |
    +-----------------------------------------------+
```

Unweighted

```
mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 1, s = 0, n = 0)
                  1               2               3
    +-----------------------------------------------+
  1 |  -.3437575758     .492302184      .238453276  |
    +-----------------------------------------------+
```

Simple weight

```
mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 1, s = 0, n = 1)
                  1               2               3
    +-----------------------------------------------+
  1 |  -.3437575758     .492302184      .238453276  |
    +-----------------------------------------------+
```

Normalized weight

```
mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 1, s = 1, n = 0)
                  1               2               3
    +-----------------------------------------------+
  1 |  -.3437575758     .492302184      .238453276  |
    +-----------------------------------------------+
```

Stabilized weight

```
mata: causal(150, 45, 20, 5, 300, 10, 40, 55, w = 1, s = 1, n = 1)
                  1               2               3
    +-----------------------------------------------+
  1 |  -.3437575758     .492302184      .238453276  |
    +-----------------------------------------------+
```

Normalized stabilized weight

Fitting MSM in Stata

// Generated simulated data with parameter = 0.3 (log hazard)

- insheet using "http://stat.ubc.ca/~e.karim/research/simdata.csv", comma

	ID	entry	exit	Outcome	tx	tx(-1)	confounder	confounder(-1)
	id	tpoint2	tpoint	y	a	am1	l	lm1
1	1	0	1	0	1	0	1	0
2	1	1	2	0	1	1	1	1
3	1	2	3	0	0	1	0	1
4	1	3	4	0	1	0	0	0
5	1	4	5	0	1	1	0	0
6	1	5	6	0	1	1	0	0
7	1	6	7	0	1	1	0	0
8	1	7	8	0	0	1	0	0
9	1	8	9	0	0	0	0	0
10	1	9	10	0	0	0	1	0
11	2	0	1	0	1	0	0	0
12	2	1	2	0	1	1	0	0
13	2	2	3	0	0	1	0	0
14	2	3	4	0	0	0	1	0
15	2	4	5	0	0	0	0	1
16	2	5	6	0	0	0	1	0
17	2	6	7	0	1	0	1	1
18	2	7	8	0	0	1	0	1
19	2	8	9	0	0	0	0	0
20	2	9	10	0	0	0	0	0

Fitting MSM in Stata

Ref: Fewell et al. (2004)

a = treatment
am1 = previous treatment
l = confounder
lm1 = previous confounder

//Calculating weights

- xi: logistic a am1 l lm1 // propensity score model for denominator
- predict pa if e(sample) // extracting fitted values
- replace pa=pa*a+(1-pa)*(1-a) // calculating probabilities for denominator
- sort id tpoint // sorting probabilities by ID
- by id: replace pa=pa*pa[_n-1] if _n!=1 // calculating cumulative probabilities

- xi: logistic a am1 // propensity score model for numerator
- predict pa0 if e(sample) // extracting fitted values
- replace pa0=pa0*a+(1-pa0)*(1-a) // calculating probabilities for numerator
- sort id tpoint // sorting probabilities by ID
- by id: replace pa0=pa0*pa0[_n-1] if _n!=1 // calculating cumulative probabilities

- gen w= 1/pa // calculating weights
- gen sw = pa0/pa // calculating stabilized weights

Fitting MSM in Stata

Ref: Fewell et al. (2004)

a = treatment
y = outcome
id = ID variable

// Simulated data parameter = 0.3 (log hazard)

//Calculating parameters from pooled logistic

- xi: logit y a, cluster(id) nolog
- xtgee y a, family(binomial) link(logit) i(id)

//Calculating parameters from pooled logistic (weighted by w)

- xi: logit y a [pw=w], cluster(id) nolog

//Calculating parameters from pooled logistic (weighted by sw)

- xi: logit y a [pw=sw], cluster(id) nolog
-

Fitting MSM in Stata

Ref: Xiao et al. (2010)

a = treatment
y = outcome
tpoint2 = entry
tpoint = exit

// Simulated data parameter = 0.3 (log hazard)
//Calculating parameters from CoxPH

- stset tpoint,   fail(y) enter(tpoint2) exit(tpoint)
- stcox a, breslow nohr

//Calculating parameters from CoxPH (weighted by w)

- stset tpoint [pw = w],   fail(y) enter(tpoint2) exit(tpoint)
- stcox a, breslow nohr

//Calculating parameters from CoxPH (weighted by sw)

- stset tpoint [pw = sw],   fail(y) enter(tpoint2) exit(tpoint)
- stcox a, breslow nohr

Fitting MSM in Stata

Using survey design setting (variable weights within same ID allowed):

- svyset id [pw = sw]
- stset tpoint , fail(y) enter(tpoint2) exit(tpoint)
- svy: stcox a, breslow nohr

Perform bootstrap to get correct standard error:

- capture program drop cboot
- program define cboot, rclass
- stcox a, breslow
- return scalar cf = _b[a]
- end
- set seed 123
- bootstrap r(cf), reps(500) cluster(id): cboot
- estat boot, all

Fitting MSM in Stata

// Simulated data parameter = 0.3 (log hazard)

//Calculating parameters from pooled logistic

y	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	.6671281	.1228866	5.43	0.000	.4262749	.9079814
_cons	-4.779052	.0926195	-51.60	0.000	-4.960583	-4.597521

//Calculating parameters from pooled logistic (weighted by w)

y	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	-.4567972	.3893786	-1.17	0.241	-1.219965	.3063709
_cons	-3.931531	.3135888	-12.54	0.000	-4.546154	-3.316908

//Calculating parameters from pooled logistic (weighted by sw)

y	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	.3180178	.147092	2.16	0.031	.0297228	.6063129
_cons	-4.575998	.1110423	-41.21	0.000	-4.793637	-4.358359

Fitting MSM in Stata

// Simulated data parameter = 0.3 (log hazard)

//Calculating parameters from CoxPH

_t	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	.6475429	.1225702	5.28	0.000	.4073097	.887776

//Calculating parameters from CoxPH (weighted by w)

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	-.4550198	.3825715	-1.19	0.234	-1.204846	.2948065

//Calculating parameters from CoxPH (weighted by sw)

_t	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
a	.3004504	.1455088	2.06	0.039	.0152584	.5856424

Simulation

Ref: Young et al. (2009)

newx = seed

subjects = number of subjects to be simulated

tpoints = number of observations per subject

// Simulation function **msm** written in mata

- **do http://stat.ubc.ca/~e.karim/research/genmsm.do**
- **mata: outputx = msm(newx = 123, subjects=2500, tpoints=10)**
- **svmat double outputx, name(outputx)**
- **renvars outputx1-outputx19 \ id tpoint tpoint2 T0 IT0 chk y ym a am1 l lm1 am1L pA_t T maxT pL psi seed**

Simulation

- Simulation Results:

	Simulation	cox	w_cox	sw_cox	logit	w_logit	sw_logit
909	909	.629	.281	.234	.638	.295	.248
910	910	.747	-.175	.259	.752	-.165	.263
911	911	.98	.209	.589	.957	.215	.564
912	912	.635	1.102	.236	.655	1.113	.252
913	913	.767	.494	.316	.764	.505	.321
914	914	.688	.188	.159	.697	.194	.17
915	915	.689	.193	.31	.697	.189	.317
916	916	.814	.625	.353	.825	.652	.365
917	917	.987	.745	.542	.98	.746	.539
918	918	.681	.267	.3	.659	.26	.278
919	919	.655	.395	.108	.636	.396	.096
920	920	.798	1.09	.4	.793	1.076	.399
921	921	.821	.251	.353	.822	.265	.352
922	922	.653	.758	.185	.641	.768	.18
923	923	.709	1.296	.272	.717	1.303	.269
924	924	.827	.678	.388	.833	.676	.395
925	925	.948	.58	.566	.959	.578	.572
926	926	.594	-.437	.108	.611	-.442	.129
927	927	.628	.29	.271	.638	.285	.275
928	928	.832	.257	.46	.836	.258	.462
929	929	.676	.674	.208	.69	.67	.222

Simulation

- Results from 1,000 simulations:

Mean of bias	No weight	W	SW
Cox	0.435	0.035	**0.008**
Logit	0.439	0.039	0.011

Median of bias	No weight	W	SW
Cox	0.438	0.040	**0.013**
Logit	0.442	0.043	**0.013**

SD	No weight	W	SW
Cox	0.118	0.412	**0.135**
Logit	0.120	0.417	**0.135**

IQR	No weight	W	SW
Cox	0.160	0.557	**0.180**
Logit	0.168	0.569	0.181

Stata vs. SAS/R

Ref: Cerdá et al. (2010)
R package: ipw

Fitting procedure

- SAS: Proc logistic for weight estimation and Proc Genmod for MSM
- R: survival package –

coxph(Surv(start, stop, event) ~ tx + cluster(id), data, weights)

- Stata: logit or stcox

Data generation (**msm** function in Mata):

- SAS/IML and R function written in the same fashion as Mata.

Acknowledgement

Joint work with:

- Dr. Paul Gustafson
- Dr. John Petkau

Department of Statistics, University of British Columbia

endMS Research and Training Network

CIHR IRSC
Canadian Institutes of Health Research
Instituts de recherche en santé du Canada

MS® Multiple Sclerosis Society of Canada

- Statalist users, special thanks to Steve Samuels

Stata® Statalist: The Stata Listserver

References

1. Robins ,J.M., Hernán, M., Brumback B. (2000). Marginal structural models and causal inference in epidemiology. Epidemiology, 11(5):550-560. [link]

2. Hernán, M., Brumback, B., and Robins, J.M. (2002). Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. Epidemiology , 11(5):561-570. [link]

3. Fewell, Z., Hernán, M., Wolfe, F., Tilling, K., Choi, H., and Sterne, J. (2004). Control-ling for time-dependent confounding using marginal structural models. Stata Journal , 4(4):402-420. [link]

4. Cerdá, M., Diez-Roux, A.V., Tchetgen Tchetgen, E., Gordon-Larsen, P., Kiefe, C. (2010) The relationship between neighborhood poverty and alcohol use: Estimation by marginal structural models, Epidemiology, 21 (4), 482-489. [link]

5. Young, J.G., Hernán, M.A., Picciotto, S., Robins, J.M. (2009) Relation between three classes of structural models for the effect of a time-varying exposure on survival. Lifetime Data Analysis, 16(1):71-84. [link]

6. Xiao, Y., Abrahamowicz, M., Moodie, E.E.M. (2010) Accuracy of conventional and marginal structural Cox model estimators: A simulation study, International Journal of Biostatistics, 6(2), 1-28. [link]

send comments to

ehsan@stat.ubc.ca

Thank You!

send comments to

ehsan@stat.ubc.ca