Handout

Outline of Handout

- Variables used for conducting survival analysis with Add Health data
- Constructing variables for Life Table Analysis and Cox regression
  - Timing of first marriage
  - Age at Wave I interview
  - High school graduation
- Constructing time-varying variables for discrete-time analysis
  - Status of first marriage
  - Status of high school graduation
- Stata commands for survival analysis
  - Life Table analysis
  - Cox regression without time-varying variables
  - Discrete-time analysis without time-varying variables
  - Discrete-time analysis with time-varying variables

Variables used for Conducting Survival Analysis with Add Health data

Table 1. Variables used for Conducting Survival Analysis with Add Health Data

Constructs	Original Variable Name	New variable Name	Response categories
First marriage and its covariates			
Marriage Sequence Number	h3mr_m_a	mar1	1st marriage[1]; legitimate skip [7]
Female	bio_sex3	female	male [1]; female [2]
Being Black	H3OD4B	black	marked [1]; Not marked[0]
Age at Wave I	N/A	N/A	N/A
Expectation of marriage at Wave I	h1ee13	expect	almost no chance [1]; some chance, but probably not [2]; a 50-50 chance [3]; a good chance [4]; almost certain [5];
High School Diploma	H3ED3	hsd	Marked [1]; Not marked[0]
GED or high school equivalency degree	H3ED2	ged	Marked [1]; Not marked[0]

BGSU Center for Family and Demographic Research

Table 1. Variables used for Conducting Survival Analysis with Add Health Data (continued)

Constructs	Original Variable Names	New variable Name	Response categories
Variables related with timing			
Year of Interview at Wave I	iyear	interview_y1	From 1994 to 1995
Month of Interview at Wave I	imonth	interview_m1	From January to December
Year of Interview at Wave III	iyear3	interview_y3	From 2001 to 2002
Month of Interview at Wave III	imonth3	interview_m3	From January to December
Year of getting married for the first time	h3mr2m_a	form_y	From 1985 to 2002
Month of getting married for the first time	h3mr2y_a	form_m	From January to December
Year of Birth from Interview at Wave I	h1gi1y	birth_y1	From 1974 to 1983
Month of Birth from Interview at Wave I	h1gi1m	birth_m1	From January to December
Year of Birth from Interview at Wave III	h3od1y	birth_y3	From 1974 to 1983
Month of Birth from Interview at Wave III	h3od1m	birth_m1	From January to December
Year of getting High School Diploma	h3ed13y	hsd_y	From 1977 to 2002
Month of getting High School Diploma	h3ed13m	hsd_m	From January to December
Year of getting General Education Diploma	h3ed12y	ged_y	From 1977 to 2002
Month of getting General Education Diploma	h3ed12m	ged_m	From January to December

Table 1. Variables used for Conducting Survival Analysis with Add Health Data (continued)

Constructs	Original Variable Names	New variable Name	Response categories
Weights variables			
Weights for region at W1	region1		West [1]; Midwest [2]; South [3]; Northeast [4];
Weights for school at WI	psuscid1		Range 001 to 371
Grand Sample Weight - W1	gswgt1		Range 16.3183 to 6649.3618
Constructed variables			
Life Tables and Cox regression			
Timing of first marriage		time	Range 1-97
Age at Wave I interview		age_t1	Range 137 to 256
High school graduation in the data for Life Tables and Cox regression		gra	Graduated [1]; Did not graduate [0]
Timing of high school graduation		gra_tm	Range 1 to 84
Indicator of being included in the analytic sample		sub_pop	Being included [1]; Not being included [0]
Discrete-time analysis			
Status of first marriage in Discrete-time analysis		married	Graduated [1]; Did not graduate [0]
High school graduation in the data for Discrete Time analysis		graduated	Graduated [1]; Did not graduate [0]

Constructing variables for Life Table analysis and Cox regression

A. Timing of first marriage or being censored

- **Create a variable for timing of first marriage**

```
gen marriage_t1 = ym(form_y1, form_m1)
label variable  marriage_t1 "century month
for getting married for the first time“
```

- **Create a variable for timing of Wave I interview**

```
gen interview_t1 = ym(iyear, imonth)
label variable interview_t1 "time for t1 interview"
```

- **Create a variable for timing of Wave III interview**

```
gen interview_t3 = ym(iyear3, imonth3)
label variable interview_t3 "time for t3 interview“
```

- **Calculate the number of months to first marriage since Wave I interview**

```
gen time1 = marriage_t1 - interview_t1  if (marriage_t1 ~=. & interview_t1~=.)
label variable time1 "time for those got married“
```

- **Calculate the number of months between Wave I and Wave III interviews**

```
gen time2 = interview_t3-interview_t1
label variable time2 "time for those did not get married“
```

- **Calculate the number of months to first marriage or being censored**

```
gen time =.
label variable time "timing of the first marriage“
replace time = time1 if time1 ~=. & mar1 ==1
replace time = time2 if mar1 ==0
replace time =. if time1 <0
```

B. Age at Wave I Interview

- **Create a birth-year variable. If an adolescent did not report his/her birth year at Wave I interview, his/her report of birth-year at Wave III interview was used.**

```
gen birth_y = birth_y1
replace birth_y = birth_y3 if birth_y1 ==.
```

- **Create a birth-month variable. If an adolescent did not report his/her birth year at Wave I interview, his/her report of the birth-month data at Wave III interview was used.**

```
gen birth_m = birth_m1
replace birth_m = birth_m3 if birth_y1 ==.
```

- **Create the month indicator for age at Wave I**

```
gen birth_t1 = ym(birth_y, birth_m)
```

- **Create the age variable**

```
gen age_t1 = interview_t1 - birth_t1
label variable age_t1 “age at WI interview”
```

C. High School Graduation

C1. Status of high school graduation

- **Create an dummy variable to indicate the status of high school graduation.**

```
gen gra =.
label variable gra "have high school or GED diplomas"
label define gra 1 "yes" 0 "no"
label value gra gra
```

- **Adolescents are viewed as having graduated from high school if they had either high school diploma or general education diploma.**

```
replace gra = 1 if (hsd ==1 | ged ==1) & mar1 ~=.
replace gra = 0 if (hsd ==0 & ged ==0) & mar1 ~=.
replace gra = 0 if (hsd ==0 & ged ==.) & mar1 ~=.
replace gra = 0 if (hsd ==. & ged ==0) & mar1 ~=.
```

C2.Timing of high school graduation

- **Create the date indicator for the time when adolescent graduated from high school or being censored**

```
gen gra_t1 = ym(gra_y, gra_m)
label variable gra_t1 "time of graduating from high school“
```

- **Calculate the timing of high school graduation**

```
gen gra_tm = gra_t1 - interview_t1  if (gra_t1 ~=. & interview_t1~=.)
label variable gra_tm “time to high school graduation”
```

- **For respondents who graduated before Wave I interview, the value of gra_tm was set to 0.**

```
replace gra_tm = 0 if (gra_t1 < interview_t1) &
(gra_t1 ~=. & interview_t1 ~=.)
```

- **For respondents who graduated after Wave III interview, the value of gra_tm was set to the months between Wave I and III interviews**

```
replace gra_tm = interview_t3 - interview_t1 if (gra_t1 > interview_t3)
& (gra_t1 ~=. & interview_t3 ~=.)
```

BGSU Center for Family and Demographic Research

Constructing time-varying variables for discrete-time analysis

A. the Status of first marriage

- **Expand the data, based on the timing of the first marriage. This step creates the person-period data for discrete-time analysis**

```
expand time
```

- **Sort the data, based on the ID variable. Create a variable, month, to indicate the number of month to which each personal record refers.**

```
sort aid
by aid: gen month=_n
```

- **Generate a variable, married, to indicate the status of first marriage. The value of the status of first marriage was initially set to 0, and this value changed to the value of marital status (i.e., mar1) if person-period records whose time indicators (i.e., month) have the same value as the variable for timing of first marriage (i.e., time).**

```
gen married=0
replace married=mar1 if month==time
```

- **Check the accuracy of the status of first marriage**

```
tab1 married if sub_pop ==1
```

B. Status of High School Graduation

- **Expand the data, based on the timing of the first marriage. This step creates the person-period data for discrete-time analysis**

```
expand time
```

- **Sort the data, based on the ID variable. Create a variable, month, to indicate the number of month to which each personal record refers.**

```
sort aid
by aid: gen month=_n
```

- **Generate a variable, graduated, to indicate the status of high school graduation. The value of the status of high school graduation was initially set to 0, and this value changes to the value of high graduation status (i.e., gra)for person-period records whose time indicators (i.e., month) have a values equals to or is grater than the value of the timing variable for high graduation (i.e., gra_m)**

```
gen graduated=0
replace graduated = gra if month >= gra_tm
```

- **Check the accuracy of high school graduation**

```
sort aid
by aid: gen N=_N
tab1 gra if supop==1 & N==month
```

Stata command for survival analysis

A. Life Table analysis

```
ltable time mar1 if sub_pop ==1, hazard

ltable time mar1 if sub_pop ==1
```

B. Cox Regression without time-varying variables

```
svyset psuscid1 [pweight = gswgt1], strata(region1)

stset time, f(mar1)

svy, subpop(sub_pop): stcox female black age_t1 expect
```

C. Discrete-time without time-varying variables

```
svyset psuscid1 [pweight = gswgt1], strata(region1)

char month [omit] 77

xi: svy, subpop(sub_pop): logistic married i.month female
black age_t1 expect
```

D. Discrete-time with time-varying variables

```
svyset psuscid1 [pweight = gswgt1], strata(region1)

char month [omit] 77

xi: svy, subpop(sub_pop): logistic married i.month female
black age_t1 expect graduated
```