

Event history analysis

An introduction

Alessandro Di Nallo

Max Planck Institute for Demographic Research

Université de Lausanne, 11 February 2025

PART II

Estimation using STATA

Estimation using STATA

- The aim of this workshop is to illustrate how to use Stata to estimate multivariate continuous & discrete time survival time models.
- These include the parametric models (e.g., Weibull hazard and piecewise constant exponential hazard functions).
- Stata provides an extensive suite of estimators. Parametric regression survival-time models are estimated by maximum likelihood

Estimation using STATA

We will take a few steps

- 1 Prepare the data for survival analysis
- 2 Understand the 'survival dynamics' of the phenomenon under investigation (Kaplan-Meier estimates)
- 3 Estimate the phenomenon with a continuous/discrete time model and time-invariant explanatory variables
- 4 Include time-varying explanatory variables

Continuous time models in STATA

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'
 - ▶ **No** complications arising from **left censoring**, **gaps**, **left truncation** ('delayed entry'), or **multiple events**

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'
 - ▶ **No** complications arising from **left censoring**, **gaps**, **left truncation** ('delayed entry'), or **multiple events**
 - ▶ There are **no missing values** for simplicity's sake

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'
 - ▶ **No** complications arising from **left censoring**, **gaps**, **left truncation** ('delayed entry'), or **multiple events**
 - ▶ There are **no missing values** for simplicity's sake
 - ▶ Data do not need to be weighted

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'
 - ▶ **No** complications arising from **left censoring**, **gaps**, **left truncation** ('delayed entry'), or **multiple events**
 - ▶ There are **no missing values** for simplicity's sake
 - ▶ Data do not need to be weighted
 - ▶ **Survival time** and **censoring variables** already exist (we have already derived them from start and end dates)

Preparing survival continuous time data for analysis

- We use the STATA suite `st` (survival time)
- `st` is very convenient for **continuous survival** time data
 - ▶ Estimation of discrete time hazard models is typically done outside this framework
- Starting assumptions
 - ▶ Data about time-to-absorbing-event data, i.e. **single failure data**
 - ▶ A single record per 'subject'
 - ▶ **No** complications arising from **left censoring**, **gaps**, **left truncation** ('delayed entry'), or **multiple events**
 - ▶ There are **no missing values** for simplicity's sake
 - ▶ Data do not need to be weighted
 - ▶ **Survival time** and **censoring variables** already exist (we have already derived them from start and end dates)
 - ▶ **No time-varying covariates**, for the moment.
- Here I assume that there is a Stata format **data set ready to use**.

Data structure of continuous time models (an example)

```
. use cancer,clear
(Patient Survival in Drug Trial)
. de
Contains data from cancer.dta
obs:          48          Patient Survival in Drug Trial
vars:          4          16 Nov 1998 11:49
size:         576 (100.0% of memory free)
```

1. studytim	int	%8.0g	Months to death or end of exp.
2. died	int	%8.0g	1 if patient died
3. drug	int	%8.0g	Drug type (1=placebo)
4. age	int	%8.0g	Patient's age at start of exp.

- studytim is the survival time
- died is the censoring indicator;
- drug and age are potential explanatory variables.

Data structure of continuous time models (an example)

studytime	died	drug	age	_st	_d	_t	_t0
11	Yes	Placebo	55	1	1	11	0
12	Yes	Placebo	49	1	1	12	0
12	Yes	Placebo	62	1	1	12	0
15	Yes	Placebo	51	1	1	15	0
17	Yes	Placebo	49	1	1	17	0
22	Yes	Placebo	57	1	1	22	0
23	Yes	Placebo	52	1	1	23	0
6	Yes	Other	67	1	1	6	0
6	No	Other	65	1	0	6	0
7	Yes	Other	58	1	1	7	0
9	No	Other	56	1	0	9	0
10	No	Other	49	1	0	10	0
11	No	Other	61	1	0	11	0

- `studytim` is the survival time
- `died` is the censoring indicator;
- `drug` and `age` are potential explanatory variables.

Data structure of continuous time models (an example)

	studytime	died	drug	age	_st	_d	_t	_t0
	11	Yes	Placebo	55	1	1	11	0
	12	Yes	Placebo	49	1	1	12	0
	12	Yes	Placebo	62	1	1	12	0
	15	Yes	Placebo	51	1	1	15	0
	17	Yes	Placebo	49	1	1	17	0
	22	Yes	Placebo	57	1	1	22	0
	23	Yes	Placebo	52	1	1	23	0
	6	Yes	Other	67	1	1	6	0
	6	No	Other	65	1	0	6	0
	7	Yes	Other	58	1	1	7	0
	9	No	Other	56	1	0	9	0
	10	No	Other	49	1	0	10	0
	11	No	Other	61	1	0	11	0

- One row in the data set for each 'subject' (e.g. person or firm).
- Columns in the data set (variables) contain at least two types of information for each subject:
 - 1 the length of time in the state (the **survival time** = the length of time the subject was exposed to the risk of experiencing a 'failure');
 - 2 **censoring status** (a variable typically equal to 1 if the person experienced a 'failure', and equal to 0 otherwise).
 - 3 At least one (time-invariant, for now) variable used as **regressor** in estimation of multivariate hazard models.

Data structure of continuous time models (an example)

`stset` creates a set of new variables in the data

```
. de *
```

variable name	storage type	display format	value label	variable label
<code>_st</code>	byte	%8.0g		
<code>_d</code>	byte	%8.0g		
<code>_t</code>	byte	%10.0g		
<code>_t0</code>	byte	%10.0g		

```
. su *
```

Variable	Obs	Mean	Std. Dev.	Min	Max
<code>_st</code>	48	1	0	1	1
<code>_d</code>	48	.6458333	.4833211	0	1
<code>_t</code>	48	15.5	10.25629	1	39
<code>_t0</code>	48	0	0	0	0

- `_st` is a 0/1 variable, equal to 1 for observations whose data has been `stset`;
- `_d` variable is the binary censoring indicator (died in this case);
- `_t` is the duration variable (`studytim` in this case);
- `_t0` records the date of entry to the study for each case, i.e. when they were first observed at risk of the event (i.e. $t = 0$);

Data structure of continuous time models (an example)

stsum shows how the data are currently set

```
. stsum
```

```
      failure _d:  died  
analysis time _t:  studytim
```

		incidence	no. of	Survival time			
	time at risk	rate	subjects	25%	50%	75%	
total		744	.0416667	48	8	17	33

- There are 31 failures (deaths in this case);
- The incidence rate, $0.04167 = 31/744$;
- The median survival time since the start of the study is 17 months

Estimation using `streg`

The basic syntax is

```
streg [varlist], dist(distname) nohr time tr nolog
```

- specifies the survival model to be estimated;
- is one of the following: exponential, **weibull**, gompertz, lognormal, loglogistic or gamma

Estimation of a Weibull model using `streg`

```
. streg drug age, dist(weibull) nolog nohr
```

```
      failure _d:  died  
analysis time _t:  studytim
```

Weibull regression -- log relative-hazard form

```
No. of subjects =          48          Number of obs   =          48  
No. of failures =          31  
Time at risk   =          744  
  
Log likelihood =  -42.931335          LR chi2(2)       =          35.39  
                                          Prob > chi2      =          0.0000
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	-2.196936	.4087791	-5.374	0.000	-2.998129	-1.395744
age	.1202027	.0371599	3.235	0.001	.0473707	.1930348
_cons	-10.58396	2.326271	-4.550	0.000	-15.14337	-6.024553
/ln_p	.5204297	.1389037	3.747	0.000	.2481834	.792676
p	1.682751	.2337403			1.281695	2.209301
1/p	.5942651	.0825456			.452632	.7802168

The `nohr` option means that **coefficient estimates** are shown.

Estimation of a Weibull model using `streg`

```
. streg, hr
```

```
Weibull regression -- log relative-hazard form
```

```
No. of subjects =           48                Number of obs   =           48
No. of failures =           31
Time at risk    =           744
Log likelihood   =   -42.931335                LR chi2(2)       =           35.39
                                                Prob > chi2        =           0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	

drug	.1111431	.045433	-5.374	0.000	.0498803	.2476487
age	1.127725	.0419062	3.235	0.001	1.048511	1.212925

/ln_p	.5204297	.1389037	3.747	0.000	.2481834	.792676

p	1.682751	.2337403			1.281695	2.209301
1/p	.5942651	.0825456			.452632	.7802168

The `hr` option means that **hazard ratios** are shown.

Interpretation of coefficient & hazard ratio of drug

$$\beta_{drug} \simeq -2.2$$

The **coefficient estimates** indicate that those receiving the drug (drug = 1) have lower [coeff. < 0] hazard rates *ceteris paribus* (i.e. lower conditional death rates and hence longer survival times).

$$\exp(\beta_{drug}) \simeq 0.11$$

Interpretation of coefficient & hazard ratio of drug

$$\beta_{drug} \simeq -2.2$$

The **coefficient estimates** indicate that those receiving the drug (drug = 1) have lower [coeff. < 0] hazard rates *ceteris paribus* (i.e. lower conditional death rates and hence longer survival times).

$$\exp(\beta_{drug}) \simeq 0.11$$

The **hazard ratio** compares the hazard rates of two categories (in this case): those who receive the drug and those who do not



The hazard *rate* for those who received the drug is only 11% of the hazard rate for those who received the placebo.

Interpretation of coefficient & hazard ratio of age

$$\beta_{(age)} \simeq 0.12$$

The **coefficient estimates** indicate that older people (age increasing) have higher [coeff. > 0] hazard rates *ceteris paribus* (i.e. higher conditional death rates and hence shorter survival times).

$$\exp(\beta_{(age)}) \simeq 1.13$$

The **hazard ratio** "compares" the hazard rates of people with age (say) X to those with age $X + 1$



The hazard *rate* of a one year rise in age ($\Delta X = 1$) is associated with a 13% higher hazard rate.

Generalizing the interpretation of β s & hazard ratios

A **dummy** (or categorical) variable (ex. drug)

β (coeff. estimate) and hazard *ratio* contrast the hazard rate of group **G** to that of the baseline group **B**

If $\beta > 0$ or hazard ratio $> 1 \rightarrow \text{hazard (G)} > \text{hazard (B)}$

If $\beta < 0$ or hazard ratio $< 1 \rightarrow \text{hazard (G)} < \text{hazard (B)}$

A **continuous** variable (ex. age)

β (coeff. estimate) and hazard *ratio* capture the **elasticity** of the hazard with respect to a **one unit** change in the explanatory variable X

If $\beta > 0$ or hazard ratio $> 1 \Rightarrow \Delta X = 1 \rightarrow \uparrow \text{hazard}$

If $\beta < 0$ or hazard ratio $< 1 \Rightarrow \Delta X = 1 \rightarrow \downarrow \text{hazard}$

More on the interpretation of β s & hazard ratios of a continuous variable

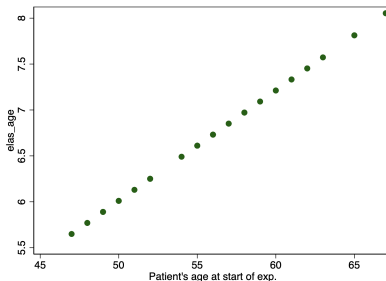
The elasticity of the hazard rate with respect to a one unit change ($\Delta X_k = 1$) in the k^{th} explanatory variable is given by $\beta_k X_{ik}$

For age, it is $0.12 * age_j$.

Let age_j vary and calculate the elasticity of the hazard wrt age:

```
ge elas_age = _b[age]*age
```

Let's plot 'elasticity of age' (y-axis) vs. 'age' (x-axis)



More on the interpretation of β s & hazard ratios of a continuous variable

More generally, hazard rate ratios at each survival time are related to absolute differences in characteristics (exploiting the properties of exponential functions):

$$h(t, X_1)/h(t, X_2) = \exp[\beta X_1]/\exp[\beta X_2] \equiv \exp[\beta(X_1 - X_2)]$$

Example 1. A 10-year difference in age, other things equal, is associated with a hazard rate ratio of some 3.3:

$$\frac{h(t; \text{age} = y + 10, \text{drug} = x)}{h(t; \text{age} = y, \text{drug} = x)} = \exp(\beta_{\text{age}} * 10) \approx 3.3$$

Example 2. Some one aged $y+10$ and who is receiving the drug has a hazard ratio that is 37% of some one aged y who gets the placebo:

$$\frac{h(t; \text{age} = y + 10, \text{drug} = 1)}{h(t; \text{age} = y, \text{drug} = 0)} = \exp(\beta_{\text{age}} * 10 + \beta_{\text{drug}} * 1) \approx 0.37$$

Question

Exercise How do we compute the hazard ratio of:

- 1 A 35-year-old patient that is receiving no drug
- 2 A 37-year-old patient that is receiving some drug (**baseline**)

Question

Exercise How do we compute the hazard ratio of:

- 1 A 35-year-old patient that is receiving no drug
- 2 A 37-year-old patient that is receiving some drug (**baseline**)

$$\frac{h(t; \text{age} = 35, \text{drug} = 0)}{h(t; \text{age} = 37, \text{drug} = 1)} =$$

Question

Exercise How do we compute the hazard ratio of:

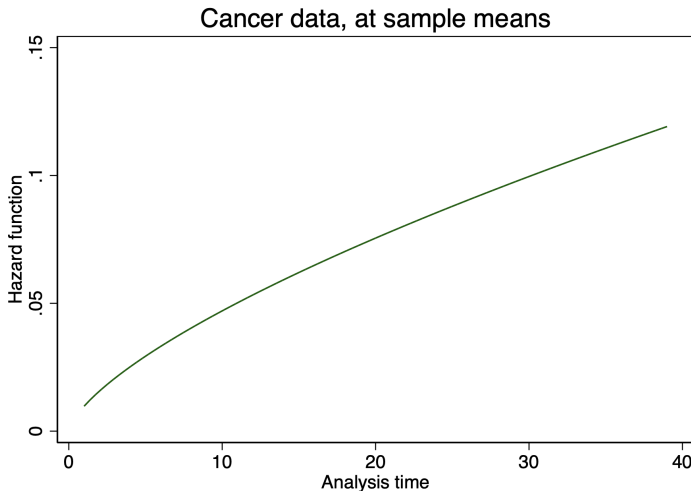
- 1 A 35-year-old patient that is receiving no drug
- 2 A 37-year-old patient that is receiving some drug (**baseline**)

$$\frac{h(t; \text{age} = 35, \text{drug} = 0)}{h(t; \text{age} = 37, \text{drug} = 1)} =$$

$$\exp[\beta_{\text{age}} * (-2) + \beta_{\text{drug}} * (-1)]$$

Estimated hazard function

```
stcurv, hazard title("Cancer data, at sample means")
```



Estimated hazard function

Recall the general formula

$$h(t; \mathbf{X}) = h_0(t) \exp(\beta \mathbf{X})$$

In this case we calculate the function $h_0(t)$ keeping the covariates at their means:

$$h(t; \overline{age}, \overline{drug}) = h_0(t) \exp(\beta \overline{\mathbf{X}})$$

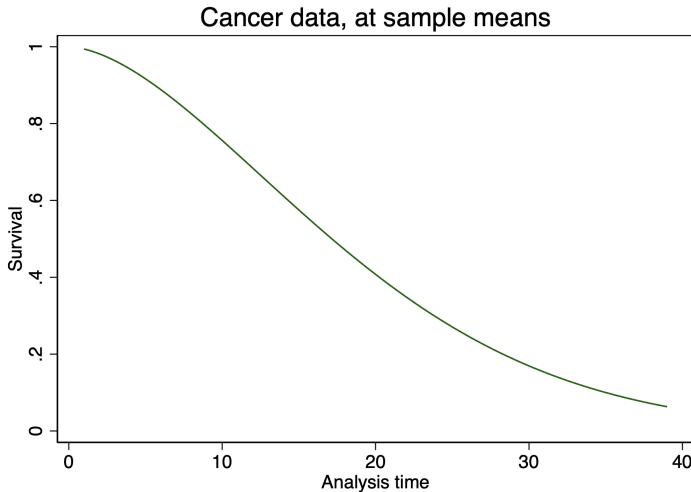
The *baseline* hazard function is a function of time, and is identical for all the individuals (or firms, etc...), regardless of their characteristics.

The curve is monotonically increasing.

The risk of death increases over time (time since treatment began).

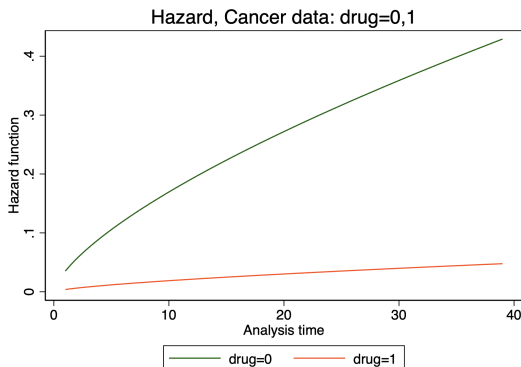
Estimated survival function

```
stcurv, survival title("Cancer data, at sample  
means")
```



Estimated hazard function

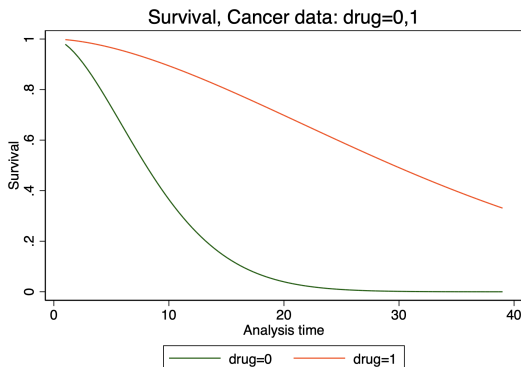
```
stcurv, hazard title("Cancer data, at sample means")
```



- Patients with the placebo ($\text{drug} = 0$): $h_0(t) \exp(\beta_{\text{age}} \overline{\text{age}})$
- Patients with the treatment ($\text{drug} = 1$): $h_0(t) \exp(\beta_{\text{age}} \overline{\text{age}} + \beta_{\text{drug}})$

Estimated survival function

```
stcurv, hazard title("Cancer data, at sample means")
```



- Patients with the placebo ($\text{drug} = 0$): $S(t, \overline{age}, \text{drug} = 0)$
- Patients with the treatment ($\text{drug} = 1$): $S(t, \overline{age}, \text{drug} = 1)$

Additional commands

`predict xb, xb`

- Generates a new variable equal to the estimate βX_i for each person i
- Calculates the shape parameter (α , in the case of Weibul) that is stored in the 'estimation class' $e(aux_p)$
- Provides the parameters to estimate, for instance, mean and median survival time for groups of people
 - ▶ Mean survival time for the 50 years old with a placebo
 - ▶ Median survival time for the 35 years old with a treatment
- More details in the lecture notes

Discrete time models in STATA

Preparing survival discrete time data for analysis and estimation

- Stata does not have a set of specialist commands for estimating the discrete time proportional odds or proportional hazards models.

Preparing survival discrete time data for analysis and estimation

- Stata does not have a set of specialist commands for estimating the discrete time proportional odds or proportional hazards models.
- All one has to do is **re-organise the data set**, define some new variables (to specify the baseline hazard function in particular), and then **apply logit or cloglog regression**.

Preparing survival discrete time data for analysis and estimation

- Stata does not have a set of specialist commands for estimating the discrete time proportional odds or proportional hazards models.
- All one has to do is **re-organise the data set**, define some new variables (to specify the baseline hazard function in particular), and then **apply logit or cloglog regression**.
 - ▶ For each unit of analysis (e.g., person), there are as many data rows as there are time intervals at risk of the event occurring for each unit of analysis.
 - ▶ "1 row - 1 person" (continuous models) \Rightarrow "Each person contributes T_i rows" (discrete models), where T_i is the number of time periods (e.g. months) i was at risk of the event.
 - ▶ A unique identifier variable for each subject is needed, plus a spell month (or semester, or year) identifier variable for each subject.

Preparing survival discrete time data for analysis and estimation

- Stata does not have a set of specialist commands for estimating the discrete time proportional odds or proportional hazards models.
- All one has to do is **re-organise the data set**, define some new variables (to specify the baseline hazard function in particular), and then **apply logit or cloglog regression**.
- The binary dependent variable also needs to be created.
 - ▶ If subject i 's survival time is **censored**, the binary dependent variable is equal to 0 for all of i 's spell months;
 - ▶ If subject i 's survival time is **not censored**, the binary dependent variable is equal to 0 for all but the last of i 's spell months (month 1,..., T_{i-1}) and equal to 1 for the last month (month T_i).
- Full tutorial on the lecture notes.

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

id[1]	1											
	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0	
13	6	2	4	Yes	0	placebo	67	1	1	4	0	
14	6	3	4	Yes	0	placebo	67	1	1	4	0	
15	6	4	4	Yes	1	placebo	67	1	1	4	0	
16	7	1	5	Yes	0	placebo	63	1	1	5	0	
17	7	2	5	Yes	0	placebo	63	1	1	5	0	
18	7	3	5	Yes	0	placebo	63	1	1	5	0	
19	7	4	5	Yes	0	placebo	63	1	1	5	0	
20	7	5	5	Yes	1	placebo	63	1	1	5	0	
21	8	1	5	Yes	0	placebo	58	1	1	5	0	
22	8	2	5	Yes	0	placebo	58	1	1	5	0	
23	8	3	5	Yes	0	placebo	58	1	1	5	0	
24	8	4	5	Yes	0	placebo	58	1	1	5	0	
25	8	5	5	Yes	1	placebo	58	1	1	5	0	
26	9	1	8	Yes	0	placebo	56	1	1	8	0	
27	9	2	8	Yes	0	placebo	56	1	1	8	0	
28	9	3	8	Yes	0	placebo	56	1	1	8	0	
29	9	4	8	Yes	0	placebo	56	1	1	8	0	
30	9	5	8	Yes	0	placebo	56	1	1	8	0	
31	9	6	8	Yes	0	placebo	56	1	1	8	0	
32	9	7	8	Yes	0	placebo	56	1	1	8	0	
33	9	8	8	Yes	1	placebo	56	1	1	8	0	
34	10	1	8	No	0	placebo	58	1	0	8	0	
35	10	2	8	No	0	placebo	58	1	0	8	0	

id: the unit of analysis (persons)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

id[1]	1										
	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

dead: period j -specific outcome

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

dead: period j -specific outcome

drug: subject to treatment (or not)

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

dead: period j -specific outcome

drug: subject to treatment (or not)

age: age when observations start

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

dead: period j -specific outcome

drug: subject to treatment (or not)

age: age when observations start

_st, _d, _t, _t0: see continuous case

Preparing survival discrete time data for analysis and estimation. An example.

Data Editor (Browse) — cancer.dta

	id	j	studytime	died	dead	drug	age	_st	_d	_t	_t0
13	6	2	4	Yes	0	placebo	67	1	1	4	0
14	6	3	4	Yes	0	placebo	67	1	1	4	0
15	6	4	4	Yes	1	placebo	67	1	1	4	0
16	7	1	5	Yes	0	placebo	63	1	1	5	0
17	7	2	5	Yes	0	placebo	63	1	1	5	0
18	7	3	5	Yes	0	placebo	63	1	1	5	0
19	7	4	5	Yes	0	placebo	63	1	1	5	0
20	7	5	5	Yes	1	placebo	63	1	1	5	0
21	8	1	5	Yes	0	placebo	58	1	1	5	0
22	8	2	5	Yes	0	placebo	58	1	1	5	0
23	8	3	5	Yes	0	placebo	58	1	1	5	0
24	8	4	5	Yes	0	placebo	58	1	1	5	0
25	8	5	5	Yes	1	placebo	58	1	1	5	0
26	9	1	8	Yes	0	placebo	56	1	1	8	0
27	9	2	8	Yes	0	placebo	56	1	1	8	0
28	9	3	8	Yes	0	placebo	56	1	1	8	0
29	9	4	8	Yes	0	placebo	56	1	1	8	0
30	9	5	8	Yes	0	placebo	56	1	1	8	0
31	9	6	8	Yes	0	placebo	56	1	1	8	0
32	9	7	8	Yes	0	placebo	56	1	1	8	0
33	9	8	8	Yes	1	placebo	56	1	1	8	0
34	10	1	8	No	0	placebo	58	1	0	8	0
35	10	2	8	No	0	placebo	58	1	0	8	0

id: the unit of analysis (persons)

j: time unit (months)

studytime: time elapsed (in months)

died: final outcome (dead or alive)

dead: period j -specific outcome

drug: subject to treatment (or not)

age: age when observations start

_st, _d, _t, _t0: see continuous case

Choose the functional form for the baseline hazard function

- We define new time-varying covariates which are functions of survival time t per person.
- Different alternatives are possible
 - ▶ $\log(\text{time})$
 - ▶ polynomial in time (e.g., $\alpha t + \gamma t^2 + \delta t^3 \dots$)
 - ▶ piece-wise constant
 - ▶ fully non-parametric (e.g., $\gamma_1 * j_1 + \gamma_2 * j_2 + \gamma_3 * j_3 + \dots + \gamma_k * j_k$, where j_k are duration-interval-specific dummy variables). *Caveat*: check whether events occur at each value of j_k .

Choose the functional form for the baseline hazard function

	id	j	studytime	lnj	j2	j3	dur1	dur2	dur3	dur4	dur5	dur6	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1	1	1	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	2	1	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	3	1	2	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4	3	2	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	4	1	3	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
6	4	2	3	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	4	3	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
8	5	1	4	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
9	5	2	4	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
10	5	3	4	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
11	5	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
12	6	1	4	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
13	6	2	4	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14	6	3	4	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
15	6	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
16	7	1	5	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
17	7	2	5	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	7	3	5	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
19	7	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
20	7	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
21	8	1	5	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
22	8	2	5	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
23	8	3	5	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
24	8	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
25	8	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
26	9	1	8	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
27	9	2	8	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
28	9	3	8	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
29	9	4	8	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
30	9	5	8	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
31	9	6	8	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
32	9	7	8	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
33	9	8	8	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
34	10	1	8	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
35	10	2	8	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
36	10	3	8	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
37	10	4	8	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
38	10	5	8	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
39	10	6	8	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
40	10	7	8	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
41	10	8	8	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
42	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Choose the functional form for the baseline hazard function

id	studytime	j	lnj	j2	j3	dur1	dur2	dur3	dur4	dur5	dur6	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1	1	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	2	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	2	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	3	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4	3	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	3	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	4	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	4	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	4	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
6	4	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
6	4	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	4	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
7	5	1	0	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	5	2	1.609438	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
7	5	3	2.079442	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7	5	4	2.79442	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
8	2	1	0	1	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
8	2	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	2	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
8	2	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
9	3	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
9	3	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
9	3	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
9	3	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	8	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	8	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
10	8	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	8	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	8	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	8	6	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
10	8	7	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
10	8	8	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0
10	8	1	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	2	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
10	8	3	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
10	8	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
10	8	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
10	8	6	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
10	8	7	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
10	8	8	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0

log(time)

Choose the functional form for the baseline hazard function

id	studytime	lnj	j	j2	j3	dur1	dur2	dur3	dur4	dur5	dur6	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1	1	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	2	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
3	2	.6931472	2	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	3	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4	3	.6931472	2	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	3	1.098612	3	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	4	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	4	.6931472	2	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	4	1.098612	3	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
5	4	1.386294	4	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6	4	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
6	4	.6931472	2	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
6	4	1.098612	3	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
6	4	1.386294	4	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
7	5	0	3	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	5	.6931472	4	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7	5	1.386294	5	25	125	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
7	5	2.079442	6	36	216	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
8	5	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	5	.6931472	2	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	5	1.386294	3	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
8	5	2.079442	4	16	64	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
9	8	0	3	9	27	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
9	8	.6931472	4	16	64	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
9	8	1.386294	5	25	125	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
9	8	2.079442	6	36	216	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
10	8	0	7	49	343	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	.6931472	8	64	512	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	1.098612	9	81	729	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	1.386294	10	100	1000	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	1.609438	11	121	1331	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	1.791759	12	144	1728	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	1.94591	13	169	2197	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	8	2.079442	14	196	2744	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Polynomial in time (e.g., $\alpha t + \gamma t^2 + \delta t^3 \dots$)

Choose the functional form for the baseline hazard function

id	j	studytime	lnj	j2	j3	dur1	dur2	dur3	dur4	dur5	dur6	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1	1	1	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	2	1	1	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	3	1	2	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	3	2	2	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	4	1	3	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	4	2	3	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	4	3	3	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	5	1	4	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	5	2	4	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
10	5	3	4	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11	5	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
12	6	1	4	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	6	2	4	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
14	6	3	4	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
15	6	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
16	7	1	5	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
17	7	2	5	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
18	7	3	5	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
19	7	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
20	7	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
21	8	1	5	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
22	8	2	5	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
23	8	3	5	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
24	8	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
25	8	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
26	9	1	8	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
27	9	2	8	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
28	9	3	8	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
29	9	4	8	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
30	9	5	8	1.609438	25	125	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
31	9	6	8	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
32	9	7	8	1.94591	49	343	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
33	9	8	8	2.079442	64	512	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
34	10	1	8	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0
35	10	2	8	.6931472	4	8	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
36	10	3	8	1.098612	9	27	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0
37	10	4	8	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
38	10	5	8	1.609438	25	125	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
39	10	6	8	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0
40	10	7	8	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
41	10	8	8	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0

Piece-wise constant

Choose the functional form for the baseline hazard function

	id	j	studytime	lnj	j2	j3	dur1	dur2	dur3	dur4	dur5	dur6	d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
1	1	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	2	1	1	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
3	3	1	2	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
4	3	2	2	.6931472	4	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	4	1	3	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
6	4	2	3	.6931472	4	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
7	4	3	3	1.098612	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
8	5	1	4	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
9	5	2	4	.6931472	4	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
10	5	3	4	1.098612	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
11	5	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
12	6	1	4	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
13	6	2	4	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
14	6	3	4	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
15	6	4	4	1.386294	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
16	7	1	5	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
17	7	2	5	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	7	3	5	1.098612	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
19	7	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
20	7	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
21	8	1	5	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
22	8	2	5	.6931472	4	8	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
23	8	3	5	1.098612	9	27	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
24	8	4	5	1.386294	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
25	8	5	5	1.609438	25	125	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
26	9	1	6	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
27	9	2	6	.6931472	4	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
28	9	3	6	1.098612	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
29	9	4	6	1.386294	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
30	9	5	6	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
31	9	6	6	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
32	9	7	6	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
33	9	8	6	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0
34	10	1	6	0	1	1	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
35	10	2	6	.6931472	4	8	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
36	10	3	6	1.098612	9	27	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
37	10	4	6	1.386294	16	64	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
38	10	5	6	1.609438	25	125	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
39	10	6	6	1.791759	36	216	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
40	10	7	6	1.94591	49	343	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
41	10	8	6	2.079442	64	512	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0

Fully non-parametric (e.g., $\gamma_1 * j_1 + \gamma_2 * j_2 + \gamma_3 * j_3 + \dots + \gamma_k * j_k$)

Estimation of a discrete-time model

Logistic model

`logit depvar varlist, [or noconstant]`

'`depvar`' is the (new) event variable – `dead` in the illustration – and '`varlist`' refers to the explanatory variables (covariates) together with the variables used to summarise the baseline hazard function.

The '`noconstant`' option means estimate a model without a constant term – we mainly use this for estimating models with a fully non-parametric baseline hazard.

Estimation of a discrete-time model

Logistic model

Odds ratios of hazard rates refer to ratios of form $\frac{h_1/(1-h_1)}{h_0/(1-h_0)}$ from a one-unit change in an explanatory variable from zero to one ($\Delta X = 1$ as $X = 0 \rightarrow X = 1$).

I personally find these difficult to interpret.

On the other hand, as $h \rightarrow 0$, the odds ratio tends to the hazard ratio $\frac{h_1}{h_0}$, which does have a ready interpretation.

Estimation of a discrete-time logistic (or logit) model

logit dead drug age lnj, nolog

Logistic regression

Number of obs = 744

LR chi2(3) = 35.41

Prob > chi2 = 0.0000

Pseudo R2 = 0.1374

Log likelihood = -111.16102

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.297714	.4388526	-5.24	0.000	-3.157849	-1.437579
age	.1253644	.0391588	3.20	0.001	.0486145	.2021143
lnj	.6781536	.2582365	2.63	0.009	.1720194	1.184288
_cons	-10.25004	2.383314	-4.30	0.000	-14.92125	-5.578826

- Drug recipients ($\text{drug} = 1$) have lower hazard rates ($\beta_{\text{drug}} < 0$ if $\text{drug} = 0 \rightarrow \text{drug} = 1$)
- Hazard rate increases with age ($\beta_{\text{age}} > 0$ if $\text{age} \uparrow$)
- Baseline hazard rises with elapsed survival time ($\ln j > 0$)

Question

Logistic regression

Log likelihood = **-111.16102**

Number of obs = **744**

LR chi2(3) = **35.41**

Prob > chi2 = **0.0000**

Pseudo R2 = **0.1374**

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.297714	.4388526	-5.24	0.000	-3.157849	-1.437579
age	.1253644	.0391588	3.20	0.001	.0486145	.2021143
lnj	.6781536	.2582365	2.63	0.009	.1720194	1.184288
_cons	-10.25004	2.383314	-4.30	0.000	-14.92125	-5.578826

What is the **hazard ratio** of a patient taking the drug vs. a patient treated with a placebo?

- ① -2.29
- ② $\exp(2.29)$
- ③ $\exp(-2.29)$
- ④ $-\exp(2.29)$

Question

Logistic regression

Log likelihood = **-111.16102**

Number of obs = **744**

LR chi2(3) = **35.41**

Prob > chi2 = **0.0000**

Pseudo R2 = **0.1374**

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.297714	.4388526	-5.24	0.000	-3.157849	-1.437579
age	.1253644	.0391588	3.20	0.001	.0486145	.2021143
lnj	.6781536	.2582365	2.63	0.009	.1720194	1.184288
_cons	-10.25004	2.383314	-4.30	0.000	-14.92125	-5.578826

What is the **hazard ratio** of a patient taking the drug vs. a patient treated with a placebo?

- ① -2.29
- ② $\exp(2.29)$
- ③ $\exp(-2.29) = 0.10$
- ④ $-\exp(2.29)$

Question

What does it mean that the **hazard ratio** of a patient taking the drug (Patient 1) vs. a patient treated with a placebo (Patient 2) is 0.10?

- ① The odds of survival for Patient 1 are 10% of Patient 2
- ② The odds of survival for Patient 2 are 10% of Patient 1
- ③ The odds of dying for Patient 2 are 90% of Patient 1
- ④ The odds of dying for Patient 1 are 10% of Patient 2

Question

What does it mean that the **hazard ratio** of a patient taking the drug (Patient 1) vs. a patient treated with a placebo (Patient 2) is 0.10?

- ① The odds of survival for Patient 1 are 10% of Patient 2
- ② The odds of survival for Patient 2 are 10% of Patient 1
- ③ The odds of dying for Patient 2 are 90% of Patient 1
- ④ The odds of dying for Patient 1 are 10% of Patient 2

Comparison of discrete and continuous-time models

Discrete time (logistic): `logit dead drug age lnj, nolog`

```
Logistic regression                                Number of obs =   744
                                                    LR chi2(3)      =   35.41
                                                    Prob > chi2     =   0.0000
Log likelihood = -111.16102                        Pseudo R2      =   0.1374
```

	dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]
drug		-2.297714	.4388526	-5.24	0.000	-3.157849 -1.437579
age		.1253644	.0391588	3.20	0.001	.0486145 .2021143
lnj		.6781536	.2582365	2.63	0.009	.1720194 1.184288
_cons		-10.25004	2.383314	-4.30	0.000	-14.92125 -5.578826

Continuous time (Weibull): `streg drug age, dist(weibull) nolog`

`nohr`

```
Weibull regression -- log relative-hazard form
```

```
No. of subjects =      48                      Number of obs   =      48
No. of failures =      31
Time at risk    =      744
Log likelihood   =  -42.931335                  LR chi2(2)      =    35.39
                                                    Prob > chi2     =    0.0000
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
drug	-2.196936	.4087791	-5.374	0.000	-2.998129 -1.395744
age	.1202027	.0371599	3.235	0.001	.0473707 .1930348
_cons	-10.58396	2.326271	-4.550	0.000	-15.14337 -6.024553

/ln_p	.5204297	.1389037	3.747	0.000	.2481834 .792676

p	1.682751	.2337403			1.281695 2.209301
1/p	.5942651	.0825456			.452632 .7802168

Comparison of discrete and continuous-time models

Discrete time (logistic): `logit dead drug age lnj, nolog`

Logistic regression

Number of obs = **744**

LR chi2(3) = **35.41**

Prob > chi2 = **0.0000**

Pseudo R2 = **0.1374**

Log likelihood = **-111.16102**

	dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]
drug		-2.297714	.4388526	-5.24	0.000	-3.157849 -1.437579
age		.1253644	.0391588	3.20	0.001	.0486145 .2021143
lnj		.6781536	.2582365	2.63	0.009	.1720194 1.184288
_cons		-10.25004	2.383314	-4.30	0.000	-14.92125 -5.578826

Continuous time (Weibull): `streg drug age, dist(weibull) nolog` `nohr`

No. of subjects = 48
No. of failures = 31
Time at risk = 744

Number of obs = **48**

LR chi2(2) = 35.39

Prob > chi2 = 0.0000

Log likelihood = -42.931335

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	-2.196936	.4087791	-5.374	0.000	-2.998129	-1.395744
age	.1202027	.0371599	3.235	0.001	.0473707	.1930348
_cons	-10.58396	2.326271	-4.550	0.000	-15.14337	-6.024553

/ln_p	.5204297	.1389037	3.747	0.000	.2481834	.792676

p	1.682751	.2337403			1.281695	2.209301
1/p	.5942651	.0825456			.452632	.7802168

Comparison of discrete and continuous-time models

Discrete time (logistic): logit dead drug age lnj, nolog

```
Logistic regression                                Number of obs =   744
                                                    LR chi2(3)      =   35.41
                                                    Prob > chi2     =   0.0000
Log likelihood = -111.16102                        Pseudo R2      =   0.1374
```

	dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug		-2.297714	.4388526	-5.24	0.000	-3.157849	-1.437579
age		.1253644	.0391588	3.20	0.001	.0486145	.2021143
lnj		.6781536	.2582365	2.63	0.009	.1720194	1.184288
_cons		-10.25004	2.383314	-4.30	0.000	-14.92125	-5.578826

Continuous time (Weibull): streg drug age, dist(weibull) nolog

nohr

```
No. of subjects =      48                                Number of obs   =      48
No. of failures =      31
Time at risk    =      744                                LR chi2(2)      =      35.39
Log likelihood   = -42.931335                            Prob > chi2     =      0.0000
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
drug	-2.196936	.4087791	-5.374	0.000	-2.998129	-1.395744
age	.1202027	.0371599	3.235	0.001	.0473707	.1930348
_cons	-10.58396	2.326271	-4.550	0.000	-15.14337	-6.024553

/ln_p	.5204297	.1389037	3.747	0.000	.2481834	.792676

p	1.682751	.2337403			1.281695	2.209301
1/p	.5942651	.0825456			.452632	.7802168

Comparison of discrete and continuous-time models

Discrete time models

Pros

- Time-varying covariates $X(t)$
- More flexible baseline hazard function

Continuous time models

Pros

- Cost-effective dataset construction
- Time-invariant covariates X
- Long time span with frequent events

Estimation of a discrete-time logit model

logistic dead drug age ln

dead	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
drug	.1004883	.0440996	-5.24	0.000	.0425171	.2375022
age	1.133561	.044389	3.20	0.001	1.049816	1.223988
lnj	1.970237	.5087869	2.63	0.009	1.187701	3.268358
_cons	.0000354	.0000843	-4.30	0.000	3.31e-07	.003777

logit dead drug age lnj, or

dead	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
drug	.1004883	.0440996	-5.24	0.000	.0425171	.2375022
age	1.133561	.044389	3.20	0.001	1.049816	1.223988
lnj	1.970237	.5087869	2.63	0.009	1.187701	3.268358
_cons	.0000354	.0000843	-4.30	0.000	3.31e-07	.003777

Logit and logistic models are equivalent.

logistic → OR

logit → coefficients (log odds)

logit [...], 'or' → OR

Estimation of a discrete-time logit model

Cubic polynomial: `logit dead drug age j j2 j3, nolog`

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.39327	.4626735	-5.17	0.000	-3.300093	-1.486447
age	.123828	.039411	3.14	0.002	.0465838	.2010721
j	.0881094	.1809724	0.49	0.626	-.26659	.4428087
j2	.0009124	.0127206	0.07	0.943	-.0240196	.0258444
j3	-.0000481	.0002508	-0.19	0.848	-.0005397	.0004436
_cons	-9.678605	2.441734	-3.96	0.000	-14.46431	-4.892895

Piece-wise constant baseline: `logit dead drug age dur1-dur6, nocons nolog`

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.280694	.4565036	-5.00	0.000	-3.175425	-1.385963
age	.1190845	.0389076	3.06	0.002	.042827	.195342
dur1	-9.098389	2.265066	-4.02	0.000	-13.53784	-4.658941
dur2	-8.434023	2.178112	-3.87	0.000	-12.70304	-4.165002
dur3	-8.438919	2.182175	-3.87	0.000	-12.7159	-4.161936
dur4	-7.596855	2.169892	-3.50	0.000	-11.84977	-3.343945
dur5	-7.445229	2.273184	-3.28	0.001	-11.90059	-2.98987
dur6	-7.499636	2.382459	-3.15	0.002	-12.16917	-2.830103

Estimation of a discrete-time logit model

Fully non-parametric baseline: logit dead drug age d1-d39, nolog

note: **d38** != 0 predicts failure perfectly;
d38 omitted and 1 obs not used.

note: **d39** != 0 predicts failure perfectly;
d39 omitted and 1 obs not used.

Logistic regression

Number of obs = 573

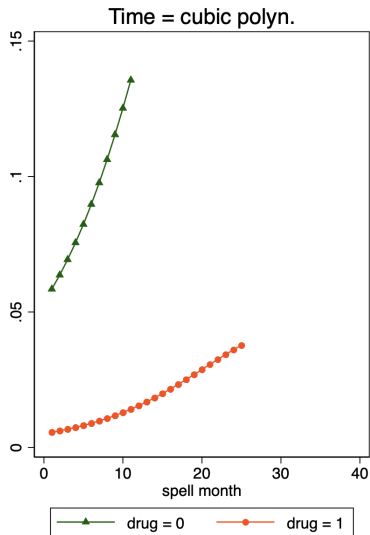
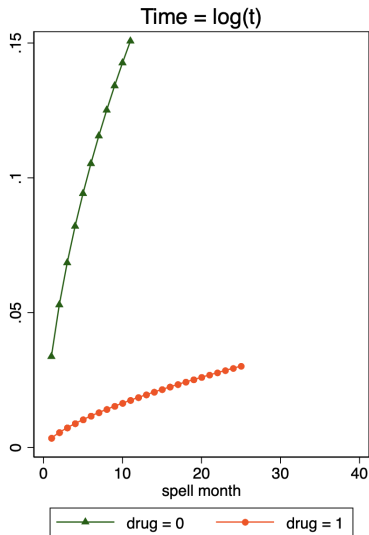
Wald chi2(23) = 159.53

Log likelihood = -96.988418

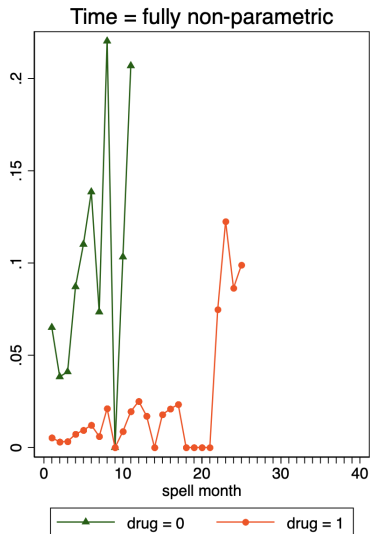
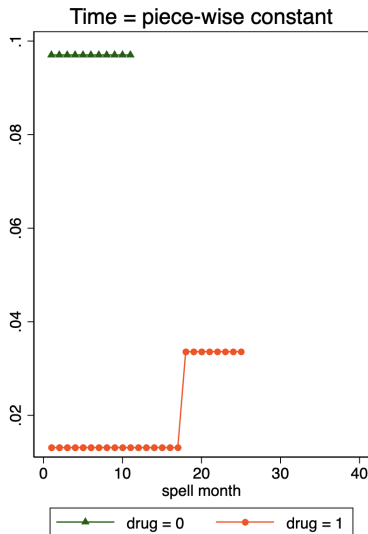
Prob > chi2 = 0.0000

dead	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
drug	-2.572486	.5104139	-5.04	0.000	-3.572879	-1.572093
age	.130051	.0406452	3.20	0.001	.0503879	.209714
d1	-9.817929	2.507439	-3.92	0.000	-14.73242	-4.903439
d2	-10.37478	2.586792	-4.01	0.000	-15.4448	-5.304765
d3	-10.30505	2.584837	-3.99	0.000	-15.37124	-5.238863
d4	-9.501784	2.473035	-3.84	0.000	-14.34884	-4.654724
d5	-9.242106	2.418842	-3.82	0.000	-13.98295	-4.501263
d6	-8.980254	2.376713	-3.78	0.000	-13.63853	-4.321983
d7	-9.687424	2.476246	-3.91	0.000	-14.54078	-4.834071
d8	-8.416641	2.31529	-3.64	0.000	-12.95453	-3.878756
d9	0 (omitted)					
d10	-9.313809	2.482374	-3.75	0.000	-14.17917	-4.448445
d11	-8.496452	2.365123	-3.59	0.000	-13.13201	-3.860896
d12	-8.242216	2.375201	-3.47	0.001	-12.89752	-3.586907
d13	-8.635718	2.447953	-3.53	0.000	-13.43362	-3.837819
d14	0 (omitted)					
d15	-8.588234	2.440998	-3.52	0.000	-13.3725	-3.803967
d16	-8.425413	2.464191	-3.42	0.001	-13.25514	-3.595688
d17	-8.312966	2.4293	-3.42	0.001	-13.0744	-3.5515

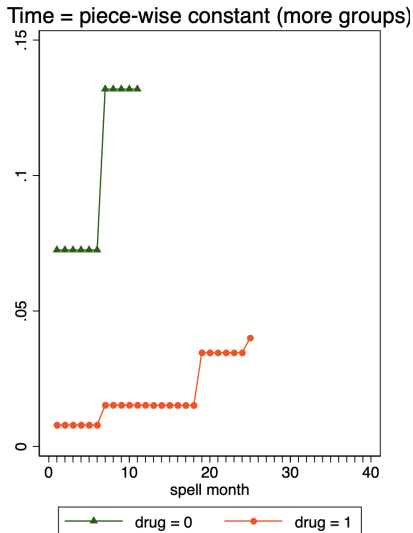
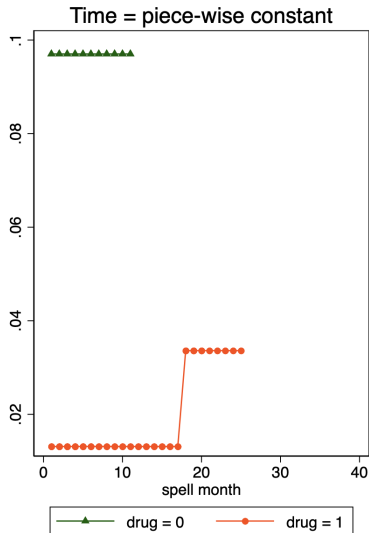
Graphical representation of a discrete-time logit hazard



Graphical representation of a discrete-time logit hazard



Graphical representation of a discrete-time logit hazard



Comparison of baseline hazard functions

Parametric functions (logarithmic, quadratic, cubic, etc.)

Pros

- Smooth hazard
- Few observations (units of analysis or transitions)

Partially or fully parametric functions (piecewise, time-dummy, etc.)

Pros

- Baseline of your choice
- Easy interpretation
- Many observations (units of analysis or transitions)

More on discrete-time models

- Parametric models
 - ▶ Complementary log-log regression
 - ▶ GLM
- Within and out-of-sample prediction
- Survival functions
- Implementation on software STATA

Full display of these topics on the Lecture notes and do-files.