

Lab 12

Survival Analysis --- two examples

These two examples come from “Modelling Survival Data in Medical Research, Second Edition” by David Collett.

Datasets on the labs website, or here:

<http://www.personal.rdg.ac.uk/~snscolet/ModellingSurvivalData/>.

Example 1.3 Survival of multiple myeloma patients.

```
use stat539_Survival_of_multiple_myeloma_patients.dta
describe
```

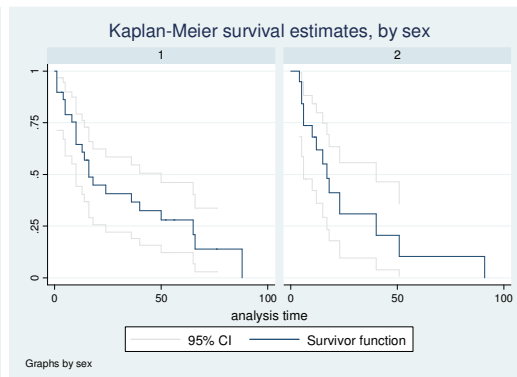
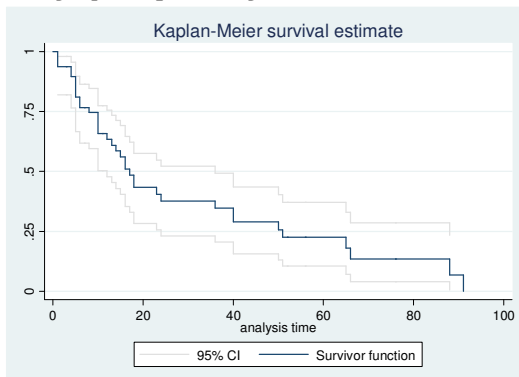
```
Contains data from
F:\USERS\Erik\UNM\public_html\courses\stat539\stat539_Survival_of_multiple_myeloma_patien
ts.dta
  obs:                48
  vars:                10                               19 Apr 2006 23:58
  size:                864 (99.9% of memory free)
```

variable name	storage type	display format	value label	variable label
patient	byte	%8.0g		
time	byte	%8.0g		
status	byte	%8.0g		
age	byte	%8.0g		
sex	byte	%8.0g		
bun	int	%8.0g		
ca	byte	%8.0g		
hb	float	%9.0g		
pcells	byte	%8.0g		
protein	byte	%8.0g		

Sorted by:

```
list, clean
```

```
* set the data as survival time data with 1 representing the event and 0 censoring
stset time, failure(status=1)
* plot the survival curve of all the data with 95% confidence bands
sts graph, gwood
* plot the survival curves by sex with 95% CBs
sts graph, by(sex) gwood
```



* these tests indicates that there isn't a difference between the sexes for experiencing the event

sts test sex

```
failure _d: status == 1
analysis time _t: time
```

Log-rank test for equality of survivor functions

sex	Events observed	Events expected
1	22	22.52
2	14	13.48
Total	36	36.00

```
chi2(1) = 0.04
Pr>chi2 = 0.8498
```

sts test sex, wilc

```
failure _d: status == 1
analysis time _t: time
```

Wilcoxon (Breslow) test for equality of survivor functions

sex	Events observed	Events expected	Sum of ranks
1	22	22.52	-1
2	14	13.48	1
Total	36	36.00	0

```
chi2(1) = 0.00
Pr>chi2 = 0.9907
```

* estimates of the hazard ratio (nearly 1) and the coefficient (nearly 0) indicate no difference between sexes

stcox sex

```
failure _d: status == 1
analysis time _t: time
```

```
Iteration 0: log likelihood = -107.96996
Iteration 1: log likelihood = -107.95282
Iteration 2: log likelihood = -107.95282
Refining estimates:
Iteration 0: log likelihood = -107.95282
```

Cox regression -- Breslow method for ties

```
No. of subjects = 48          Number of obs = 48
No. of failures = 36
Time at risk = 1122
Log likelihood = -107.95282    LR chi2(1) = 0.03
                               Prob > chi2 = 0.8531
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
sex	1.068043	.3786088	0.19	0.853	.533148 2.139586

stcox sex, nohr

failure _d: status == 1
 analysis time _t: time

Iteration 0: log likelihood = -107.96996
 Iteration 1: log likelihood = -107.95282
 Iteration 2: log likelihood = -107.95282
 Refining estimates:
 Iteration 0: log likelihood = -107.95282

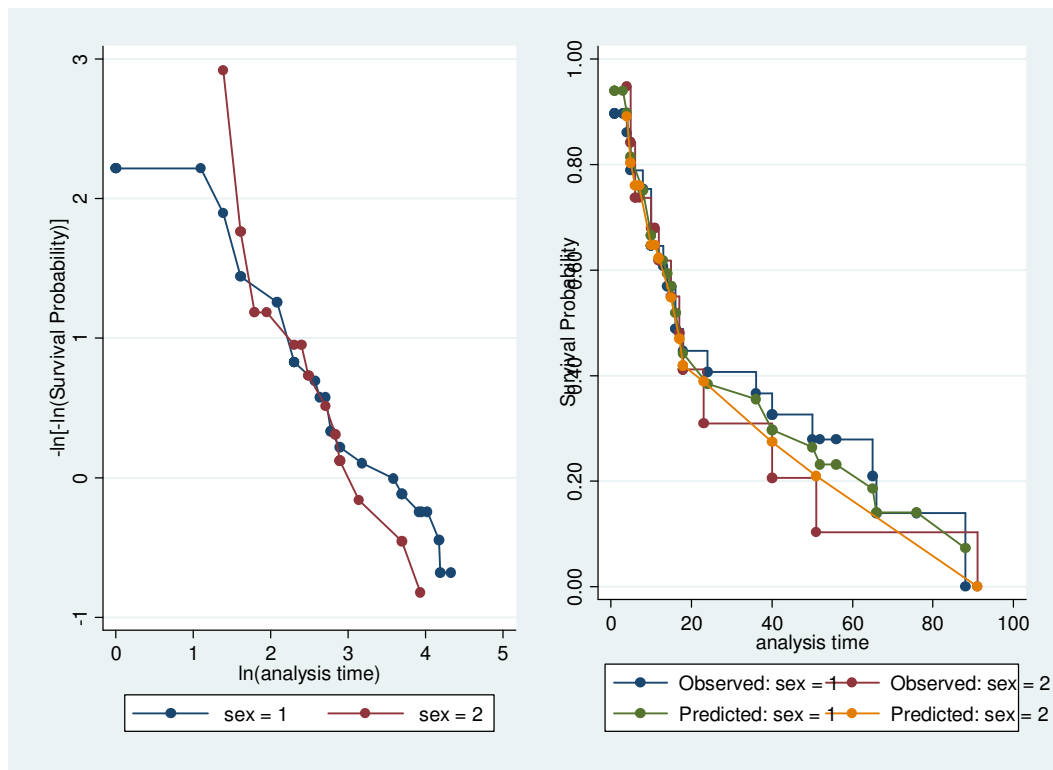
Cox regression -- Breslow method for ties

No. of subjects =	48	Number of obs =	48
No. of failures =	36		
Time at risk =	1122		
Log likelihood =	-107.95282	LR chi2(1) =	0.03
		Prob > chi2 =	0.8531

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
sex	.0658281	.3544883	0.19	0.853	-.6289563 .7606124

* the plots by sex lie nearly on top of each other

stpplot, by(sex) name(loglog)
stcoxkm, by(sex) name(phkm)
graph combine loglog phkm



* Cox Proportional Hazards Model for sex gives same result as above

xi: stcox i.sex

i.sex _Isex_1-2 (naturally coded; _Isex_1 omitted)

 failure _d: status == 1
 analysis time _t: time

Iteration 0: log likelihood = -107.96996
 Iteration 1: log likelihood = -107.95282
 Iteration 2: log likelihood = -107.95282
 Refining estimates:
 Iteration 0: log likelihood = -107.95282

Cox regression -- Breslow method for ties

No. of subjects =	48	Number of obs =	48
No. of failures =	36		
Time at risk =	1122		
Log likelihood =	-107.95282	LR chi2(1) =	0.03
		Prob > chi2 =	0.8531

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
_Isex_2	1.068043	.3786088	0.19	0.853	.533148 2.139586

* there doesn't appear to be an age effect, either
stcox age

 failure _d: status == 1
 analysis time _t: time

Iteration 0: log likelihood = -107.96996
 Iteration 1: log likelihood = -107.90871
 Iteration 2: log likelihood = -107.90871
 Refining estimates:
 Iteration 0: log likelihood = -107.90871

Cox regression -- Breslow method for ties

No. of subjects =	48	Number of obs =	48
No. of failures =	36		
Time at risk =	1122		
Log likelihood =	-107.90871	LR chi2(1) =	0.12
		Prob > chi2 =	0.7263

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
age	1.009775	.0281037	0.35	0.727	.9561683 1.066387

```

*****
* BETWEEN THESE BARS I'M TRYING TO DO A FULL MODEL, BUT COULDN'T GET IT TO WORK

* use xi to create indicator variables for a full model with all main effects and two-way
interactions
* note that the categorical i.* variables need to come first in the interactions for xi
to code them correctly
* stepwise methods should work with swcox
xi: stcox age i.sex bun ca hb pcells i.protein i.sex*age age*bun age*ca age*hb
age*pcells i.protein*age i.sex*bun i.sex*ca i.sex*hb i.sex*pcells i.sex*i.protein
bun*ca bun*hb bun*pcells i.protein*bun ca*hb ca*pcells i.protein*ca hb*pcells
hb*i.protein i.protein*pcells
describe

stcox age _Isex_2 bun ca hb pcells _Iprotein_1 _IsexXage_2 age*bun age*ca age*hb
age*pcells _IproXage_1 _IsexXbun_2 _IsexXca_2 _IsexXhb_2 _IsexXpcell_2 _IsexXpro_2_1
bun*ca bun*hb bun*pcells _IproXbun_1 ca*hb ca*pcells _IproXca_1 hb*pcells hb*i.protein
_IproXpcell_1
*****

* a main-effects model
xi: stepwise, pr(.1): stcox age i.sex bun ca hb pcells i.protein
i.sex          _Isex_1-2          (naturally coded; _Isex_1 omitted)
i.protein      _Iprotein_0-1      (naturally coded; _Iprotein_0 omitted)
              begin with full model
p = 0.9211 >= 0.1000 removing ca
p = 0.8149 >= 0.1000 removing pcells
p = 0.5294 >= 0.1000 removing _Isex_2
p = 0.5605 >= 0.1000 removing age
p = 0.1288 >= 0.1000 removing _Iprotein_1

Cox regression -- Breslow method for ties

No. of subjects =          48                Number of obs   =          48
No. of failures =          36
Time at risk    =          1122
Log likelihood  =       -101.4691
LR chi2(2)      =          13.00
Prob > chi2     =          0.0015

-----+-----
      _t | Haz. Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      bun |   1.018745   .0057807    3.27  0.001    1.007478    1.030138
      hb  |   .8749059   .0542884   -2.15  0.031    .7747183    .9880499
-----+-----

```

A person with a unit increase in BUN is 1.018745 times more likely to experience the event (death from multiple myeloma). A 100 unit increase in BUN is $1.018745^{100}=6.405$ times as likely to experience the event.

A person with a unit increase in HB is 0.8749 more likely (that is, they are less likely) to experience the event.

index	2.176623	.7876037	2.15	0.032	1.070974	4.42372
size	1.109334	.050235	2.29	0.022	1.015119	1.212295

With index and size in the model, treatment is no longer significant (p-value of 0.3550). A person with a unit increase in INDEX is 2.176623 times more likely to experience the event (death from prostatic cancer). A 3 unit increase in INDEX is $2.176623^3=10.31$ times as likely to experience the event. A person with a unit increase in SIZE is 1.109334 more likely to experience the event.

* to get nice plots, partition these continuous variables into a few groups, and compare the groups.

* here I use sex just as an example, even though we know there is no sex difference.

```
sts graph, by(treatment) name(km)
sts graph, gwood by(treatment) name(gw)
stphplot, by(treatment) name(loglog)
stcoxkm, by(treatment) name(phkm)
graph combine km gw loglog phkm
```

Kaplan-Meier survival estimates, by treatment

