

~/survival-models/doc/homework-1.txt

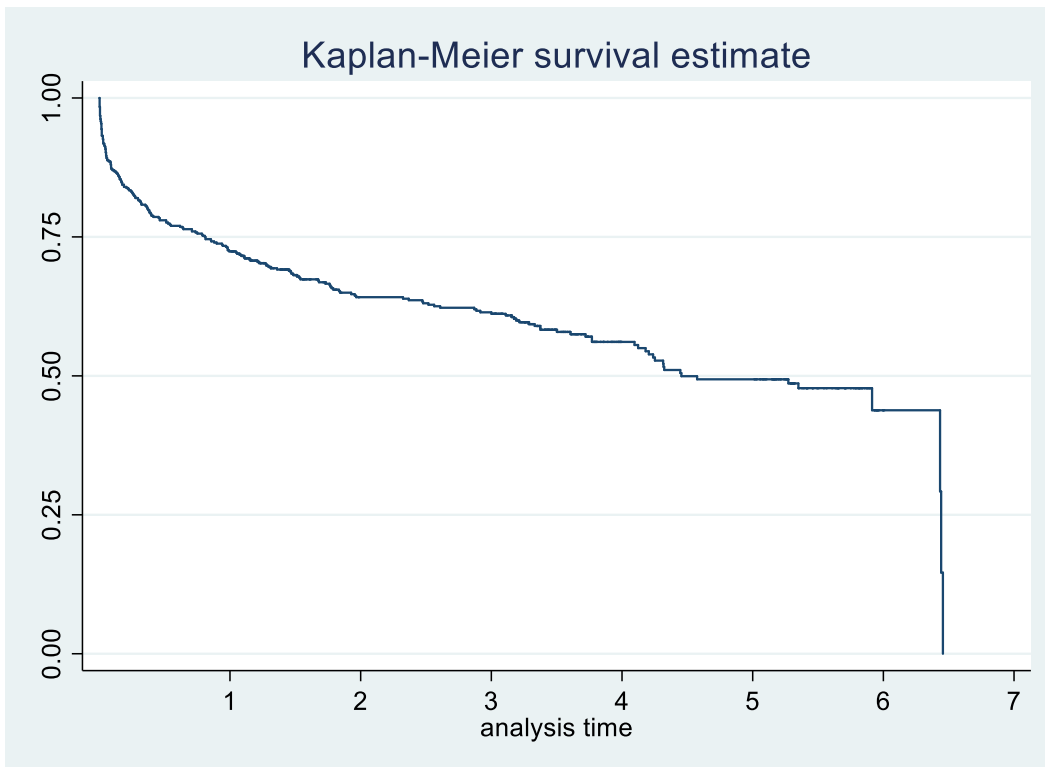
**1. Open the WHAS500 data set in the software program of your choice.**

a. Produce a table of counts for fstat, to indicate which patients have died and which have been censored.

```
. tab fstat
```

Vital Satus	Freq.	Percent	Cum.
Alive	285	57.00	57.00
Dead	215	43.00	100.00
Total	500	100.00	

b. Draw a Kaplan-Meier plot for overall survival.



c. Estimate the 25th, 50th, and 75th quantiles for overall survival.

. stci, p(25)

```
failure _d: fstat
analysis time _t: lenfol
```

	no. of subjects	25%	Std. Err.	[95% Conf. Interval]	
total	500	295	63.22275	146	406

. stci, median

```
failure _d: fstat
analysis time _t: lenfol
```

	no. of subjects	50%	Std. Err.	[95% Conf. Interval]	
total	500	1627	159.5555	1506	2353

. stci, p(75)

```
failure _d: fstat
analysis time _t: lenfol
```

	no. of subjects	75%	Std. Err.	[95% Conf. Interval]	
total	500	2353	66.28936	2350	.

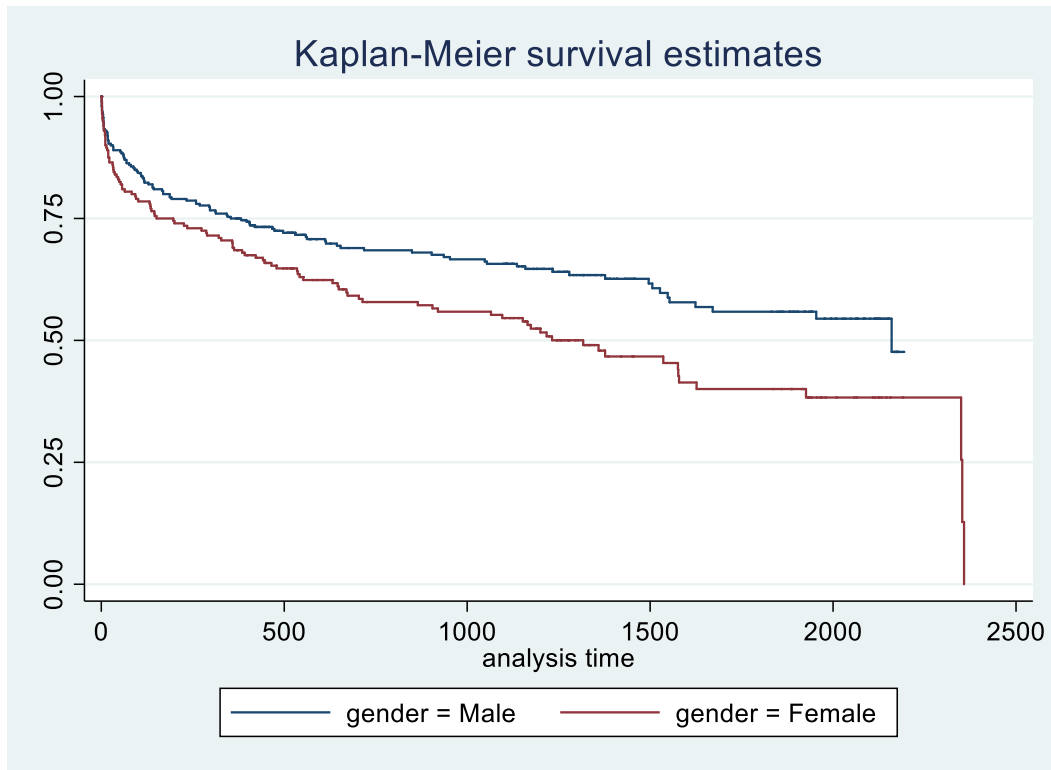
## 2. Use the WHAS500 data set for this problem.

a. Produce a crosstabulation of fstat and gender. Are you comfortable with the number of deaths in each group?

gender	Vital Satus		Total
	Alive	Dead	
Male	189	111	300
Female	96	104	200
Total	285	215	500

Pearson chi2(1) = 11.0159 Pr = 0.001

b. Draw Kaplan-Meier curves for males and females.



c. Calculate median survival with confidence intervals for males and females.

```
. stci, median by(gender)
```

```
      failure _d:  fstat
analysis time _t:  lenfol
```

gender	no. of subjects	50%	Std. Err.	[95% Conf. Interval]	
Male	300	2160	.	1624	.
Female	200	1317	177.0388	865	1579
total	500	1627	159.5555	1506	2353

d. Calculate the log rank test for males versus females. Interpret your result.

Log-rank test for equality of survivor functions

gender	Events observed	Events expected
Male	111	130.73
Female	104	84.27
Total	215	215.00

chi2(1) = 7.79  
Pr>chi2 = 0.0053

**3. Use the WHAS500 data set for this problem.**

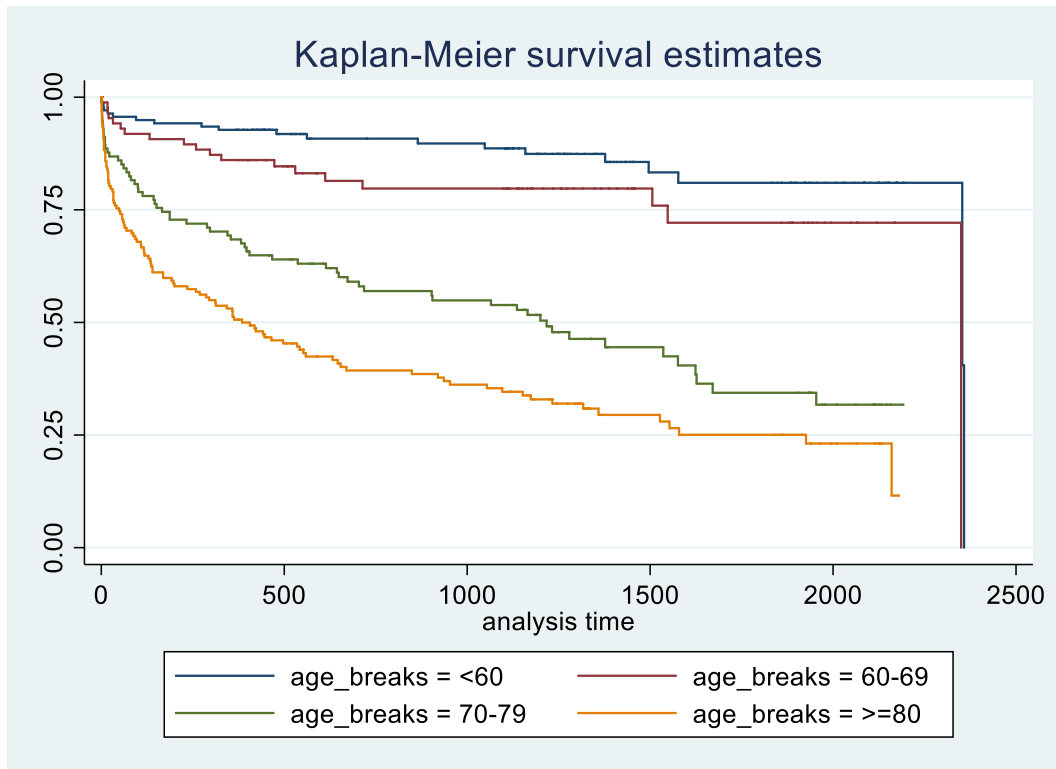
a. Produce age groups <60, 60-69, 70-79, and >=80.

Compute a crosstabulation of this variable with fstat. Are you comfortable with the number of deaths in each group?

age_breaks	Vital Satus		Total
	Alive	Dead	
<60	118	20	138
60-69	67	19	86
70-79	50	64	114
>=80	50	112	162
Total	285	215	500

Pearson chi2(3) = 114.2723 Pr = 0.000

b. Draw Kaplan Meier curves for each age group.



c. Calculate the median survival time with confidence intervals for each age group.

```
. stci, median by(age_breaks)
```

```
      failure _d:  fstat
analysis time _t:  lenfol
```

age_breaks	no. of subjects	50%	Std. Err.	[95% Conf. Interval]	
<60	138	2353	114.2134	2353	.
60-69	86	2350	16.64846	.	.
70-79	114	1217	204.2513	673	1624
>=80	162	385	89.30563	235	632
total	500	1627	159.5555	1506	2353

d. Calculate the log rank test for age groups. Interpret your results.

Log-rank test for equality of survivor functions

age_breaks	Events observed	Events expected
<60	20	71.96
60-69	19	41.34
70-79	64	47.52
>=80	112	54.18
Total	215	215.00
chi2 (3) =		120.13
Pr>chi2 =		0.0000

4. (Only for those who are brave) The following are times for catheters in infants. A "+" means that the catheter was removed because it was no longer needed. Times without a + mean that the catheter was removed because it failed. Occlusion and infection were the two major reasons for failure. Treating failures as an event and removal because it was no longer needed as a censored observation, estimate the Kaplan-Meier survival curve by hand, showing all your intermediate calculations.

1+, 1+, 1+, 1+, 1+, 1+, 1+, 1+, 1, 1, 2+, 2+, 2, 2, 3+, 3, 4+, 4, 5+, 5+, 5+, 5+, 5+, 5, 5, 5, 5, 6, 6, 7, 10, 10, 12, 12, 13

See do-file