MATH 60604A
Statistical modelling
§ 7e - Log rank test

HEC Montréal
Department of Decision Sciences

Comparing survival curves

Recall the breast cancer trial data

- `time`: time until death, or end of study in months
- `death`: death indicator, 0 for survivors, 1 for deceased
- `im`: response to immunohistochemical examination, either negative (0) or positive (1)

Suppose we're interested in seeing whether survival function is different for those who respond positively relative to those who respond negatively to the immunohistochemical examination.

- Do women who respond positive to the immunohistochemical examination tend to survive longer than those who respond negatively?

Compararison of survival curves

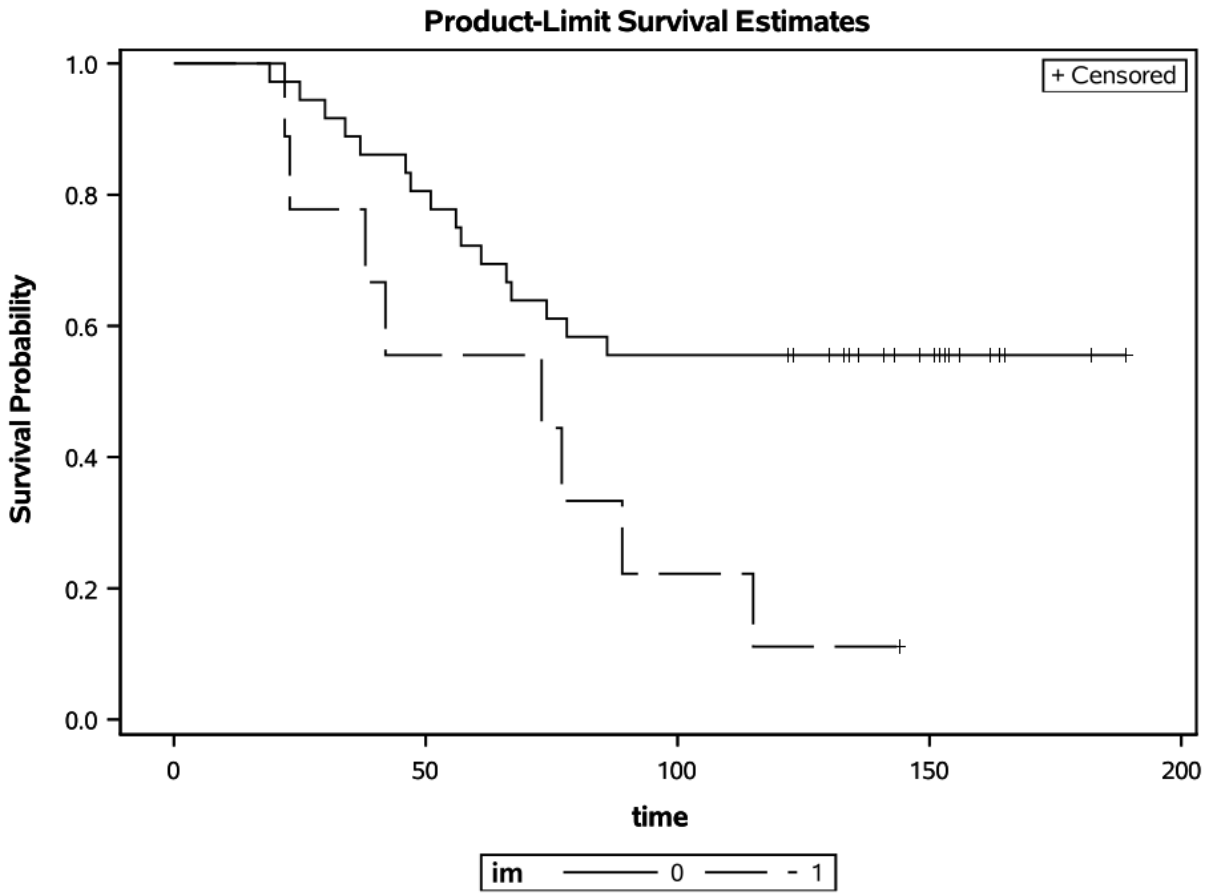
We specify the group variable through the command `strata`.

SAS

```
proc lifetest data=statmod.breastcancer method=km plots=(s);
time time*death(0);
strata im;
run;
```

SAS will estimate the survival curve separately for individuals in the two `im` groups.

Survival curves (Kaplan–Meier)



Comparison of survival curves

It seems that women with a negative immunohistochemical examination response (`im=0`) have better survival than those who have a positive response (`im=1`).

- For most times t, $\widehat{S}_1(t) > \widehat{S}_2(t)$ so that those with `im=0` have higher survival probability than those with `im=1`.

But is is the survival function significantly different in the two groups `im=0` and `im=1`?

$\mathscr{H}_0 : S_1(t) = S_2(t)$ for all t,

$\mathscr{H}_1 : S_0(t) \neq S_1(t)$ for at least one value of t.

Log rank test

Consider Cox's proportional hazard regression model of the form

$$h(t) = h_0(t)\exp(\beta \texttt{im}). \qquad (\star)$$

- The null hypothesis of equality of survival functions amounts to testing $\mathscr{H}_0 : \beta = 0$.
- The score statistic can be used to test this hypothesis without fitting the model
  - under $\mathscr{H}_0$, the estimated hazard yields Kaplan–Meier's estimator.
- The score test requires calculating the gradient and the hessian of the model $(\star)$ and evaluating them at $\beta = 0$
  - both are simple function of the number of individuals at risk each time t_i.

Proportional hazard model

SAS code to fit a proportional hazard model

```
proc phreg data=statmod.breastcancer;
model time*death(0) = im / ties=exact;
run;
```

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.4943	1	0.0191
Wilcoxon	4.3512	1	0.0370
-2Log(LR)	5.6708	1	0.0172

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.4463	1	0.0350
Score	5.4943	1	0.0191
Wald	5.0804	1	0.0242

The log rank test is also displayed by default in the SAS output of the `lifetest` procedure (left).

Log rank test

- Under $\mathscr{H}_0 : \beta = 0$, the null distribution of the score statistic is approximately χ^2_1.
- The p-value is 0.0191: we reject $\mathscr{H}_0$ at level 5% and conclude that the survival curves are different for women with negative / positive responses to the immunohistochemical examination.
- We can generalize the log rank test by using a Cox model with a k level categorical variable as sole predictor:
  - the null distribution of the test statistic is then χ^2_{k-1}.