Survival Analysis for Randomized Clinical Trials

José Sánchez 2020-02-21

Part I: Kaplan-Meier estimation

1. INTRODUCTION TO SURVIVAL TIME DATA (also known as time-to-event data)
2. ESTIMATING THE SURVIVAL FUNCTION
3. (THE LOG RANK TEST)

Survival Analysis

Important in:

- Clinical Trials (eg cancer treatments)
- Life Insurance & Superannuation
- Planning (eg aged care)
- Component Manufacture

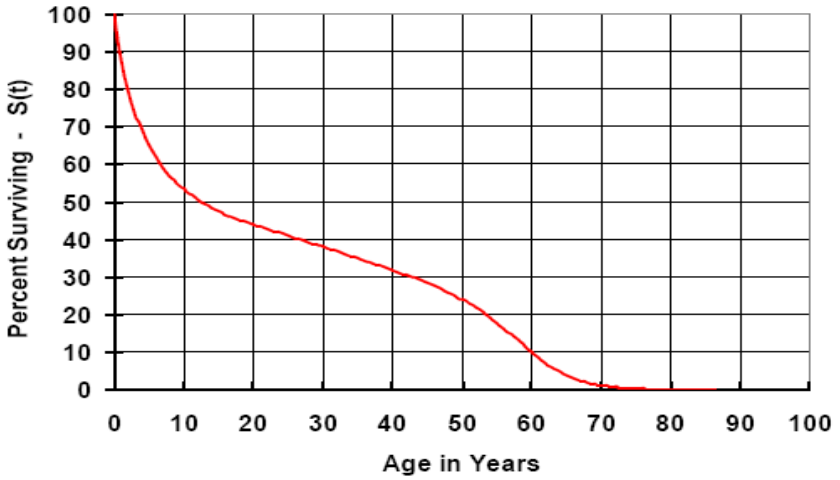
Concept: **Survival Function or Curve**

Measures the survival over time from a particular starting time (eg birth, treatment, diagnosis) to a particular endpoint (death, failure, relapse).

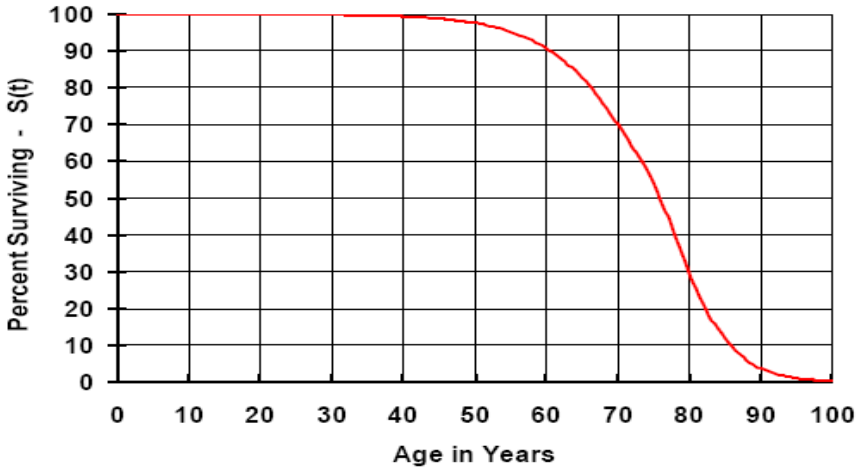
Example: $S(85) = 0.084$ (males)

$S(85) = 0.197$ (females)

Third World Countries – All Cause Mortality

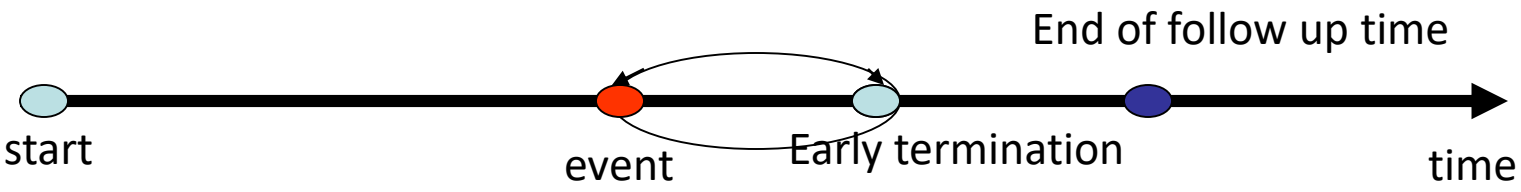


Western Countries – All Cause Mortality



Elements of Survival Experiments

- Event Definition (death, adverse events, …)
- Starting time
- Length of follow-up (equal length of follow-up, common stop time)
- Failure time (observed time of event since start of trial)
- Unobserved event time (censoring, no event recorded in the follow-up, early termination, etc)



When to use survival analysis

- Examples
  - Time to death or clinical endpoint
  - Time in remission after treatment for cancer
  - Recidivism rate after alcohol treatment
- When one believes that 1+ explanatory variable(s) explains the differences in time to an event
- Especially when follow-up is incomplete or variable

Survival Analysis in RCT

- For survival analysis, the best observation plan is prospective. In clinical investigation, that is a randomized clinical trial (RCT).
- Random treatment assignments.
- Well-defined starting points.
- Substantial follow-up time.
- Exact time records of the interesting events.

Survival Analysis in Observational Studies

- Survival analysis can be used in observational studies (cohort, case control etc) as long as you recognize its limitations.
- Lack of causal interpretation.
- Unbalanced subject characteristics.
- Determination of the starting points.
- Lost of follow-up.
- Ascertainment of event times.

Standard Notation for Survival Data

- T_i -- Survival (failure) time
- C_i -- Censoring time
- $X_i = min\,(T_i, C_i)$ -- Observed time
- $\Delta_i = I\,(T_i \leq C_i)$ -- Failure indicator: If the ith subject had an event before been censored, $\Delta_i = 1$, otherwise $\Delta_i = 0$.
- $\boldsymbol{Z}_i(t)$ – covariate vector at time t.
- Data: $\{X_i\,,\ \Delta_i\,,\ \boldsymbol{Z}_i(\cdot)\ \}$, where $i = 1,2,\ldots n$.

Describing Survival Experiments

- Central idea: the event times are realizations of an unobserved stochastic process, that can be described by a probability distribution.
- Description of a probability distribution:
  1. Cumulative distribution function, F(t)
  2. Survival function, S(t)
  3. Probability density function, f(t)
  4. Hazard function, h(t)
  5. Cumulative hazard function, H(t)

Relationships Among Different Representations

- Given any one, we can recover the others.

$$S(t) = P(T > t) = 1 - F(t) = \exp\{-\int_0^t h(u)du\}$$

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T \le t + \Delta t \mid T \ge t)}{\Delta t} = \frac{f(t)}{S(t)}$$

$$h(t) = -\frac{\partial}{\partial t} \log S(t) \quad \longrightarrow \quad H(t) = \int_0^t h(u)du$$

$$f(t) = h(t) \exp\{-\int_0^t h(u)du\}$$

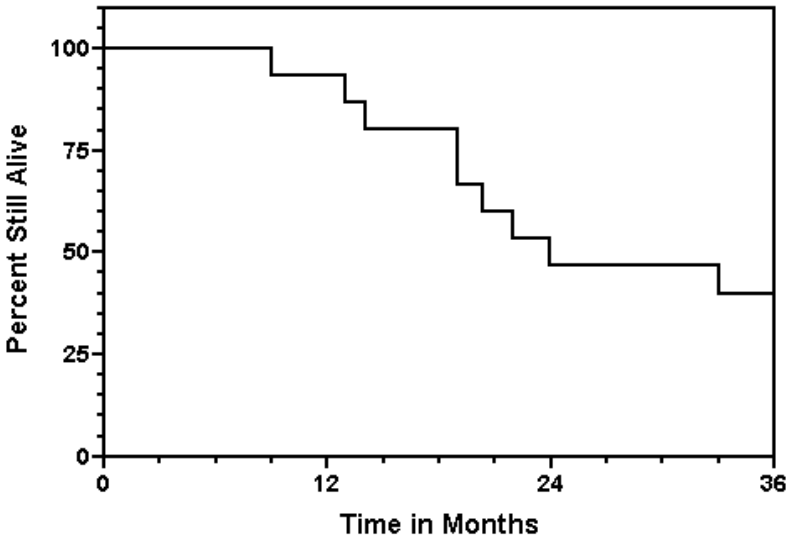
Descriptive statistics

- **Average survival**
  - Can we calculate this with censored data?
- **Average hazard rate**
  - Total # of failures divided by observed survival time (units are therefore 1/t or 1/pt-yrs)
  - An incidence rate, with a higher value indicating lower survival probability
- Provide an overall statistic only

Estimating the survival function

There are two slightly different methods to create a survival curve.

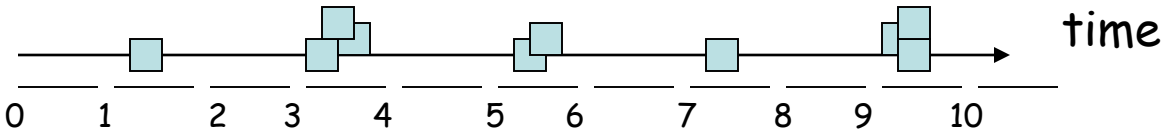
- With the actuarial method, the x axis is divided up into regular intervals, perhaps months or years, and survival is calculated for each interval.
- With the Kaplan-Meier method, survival is recalculated every time a patient dies. This method is preferred, unless the number of patients is huge.



The term life-table analysis is used inconsistently, but usually includes both methods.

Life Tables (no censoring)

In survival analysis, the object of primary interest is the survival function S(t).Therefore we need to develop methods for estimating it in a good manner. The most obvious estimate is the empirical survival function:



$$\hat{S}(t) = \frac{\#\,\text{patients with survival times larger than t}}{\text{Total}\,\#\,\text{patients}}$$

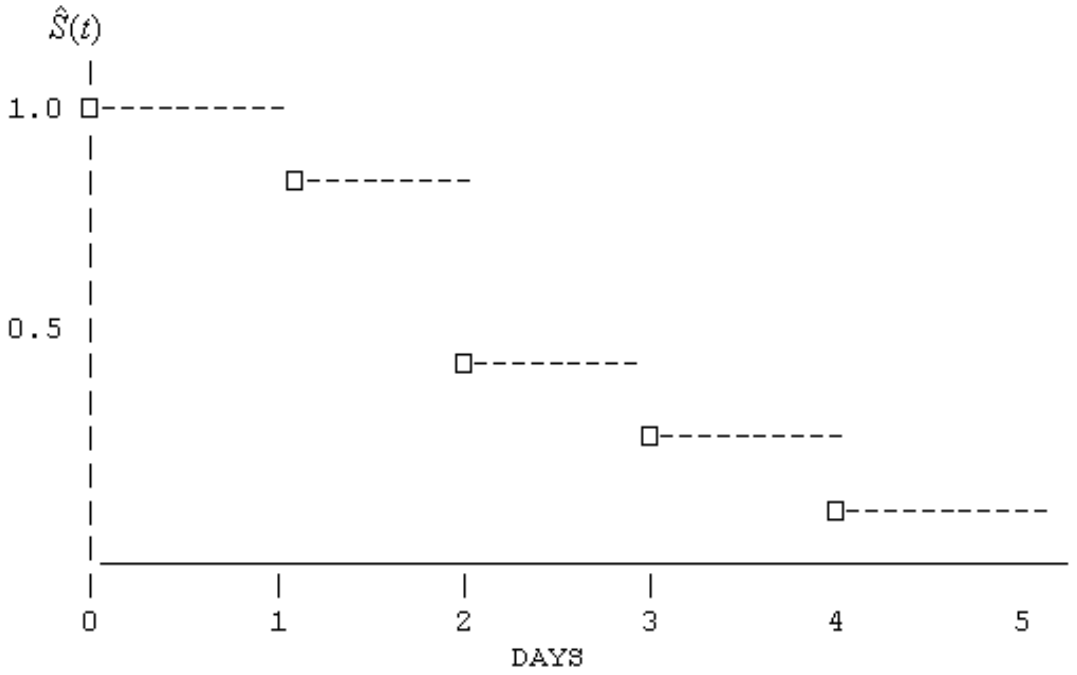
$$\hat{S}(0) = \frac{10}{10} = 1 \qquad \hat{S}(1) = \frac{10}{10} = 1 \qquad \hat{S}(2) = \frac{9}{10} = 0.9 \qquad \hat{S}(5) = \frac{6}{10} = 0.6$$

Example: a rat survival study

- In an experiment, 20 rats exposed to a particular type of radiation were followed over time.
- The **start time** of follow-up was the same for each rat. This is an important difference from clinical studies where patients are recruited into the study over time and at the date of the analysis had been followed for different lengths of time.
- In this simple experiment all individuals have the same potential follow-up time. The potential **follow-up** time for each of the 20 rats is 5 days.

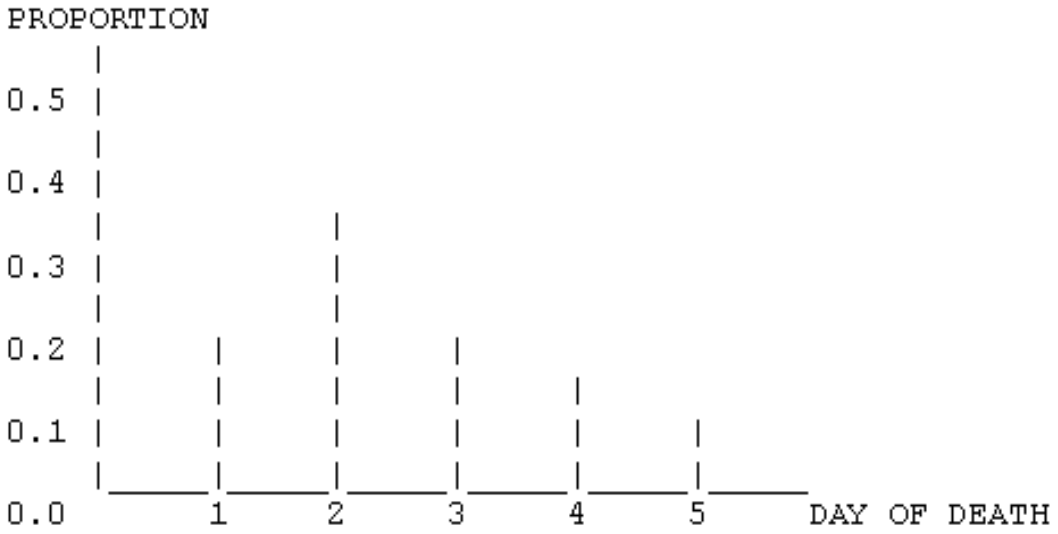
Survival Function for Rats

DAY	NUMBER ALIVE START OF DAY	NUMBER DIE AT DURING DAY	NUMBER ALIVE AT END OF DAY	PROPORTION DIE DURING DAY *** $\hat{P}[T=t]$	PROPORTION SURVIVING TO END OF DAY ### $\hat{S}(t)=\hat{P}[T\geq t]$
1	20	4	16	4/20 = 0.20	16/20 = 0.80
2	16	7	9	7/20 = 0.35	9/20 = 0.45
3	9	4	5	4/20 = 0.20	5/20 = 0.25
4	5	3	2	3/20 = 0.15	2/20 = 0.10
5	2	2	0	2/20 = 0.10	0/20 = 0.00
				1.00	



Survival Curve for Rat Study

Proportion of rats dying on each of 5 days



Confidence Intervals for Survival Probabilities

- From above we see that the "cumulative“ probability of surviving three days in the rat study is 0.25.
- We may want to report this probability along with its standard error. This sample proportion of 0.25 is based on 20 rats that started the study. If we assume that
  - (i) each rat has the same unknown probability of surviving three days, S(3), and
  - (ii) assume that the probability of one rat dying is not influenced by whether or not another rat dies,

  then we can use results associated with the **binomial probability distribution** to obtain the variance of this proportion

$$VARIANCE[\,S(3)\,] = \frac{S(3)\times[1-S(3)]}{n} = \frac{0.25\times 0.75}{20} = 0.009375$$

$$Z = \frac{\hat{S}(3) - S(3)}{\sqrt{Var\left[\hat{S}(3)\right]}} \cong N(0,1)$$

$$Var\left[\hat{S}(3)\right] \approx \frac{\hat{S}(3)\left(1-\hat{S}(3)\right)}{20}$$

- This can be used to test hypotheses about the theoretical probability of surviving three days as well as to construct confidence intervals.
- For example, the 95% confidence interval for is given by

0.25 +/- 1.96 x 0.094 or (0.060,0.440)

We are 95% confident that the probability of surviving 3 days, meaning **THREE OR MORE DAYS**, lies between 0.060 and 0.440.

This situation is not realistic. In a RCT we have that

1. Patients are recruited at different time periods
2. Some observations are censored
3. Patients can differ wrt many covariates
4. We should avoid discretising continuous data if possible

Kaplan-Meier survival curves

- Also known as **product-limit** formula
- Accounts for censoring
- Generates the characteristic “stair step” survival curves
- Does **not** account for confounding or effect modification by other covariates
  - Is that a problem?

Censored Observations (Kaplan-Meier)

We proceed as in the case without censoring

$$P_i = \Pr\left(T \geq i \mid T \geq i-1\right)$$

$$S\left(k\right) = P_1 \times P_2 \times \ldots \times P_k$$

P_i stands for the proportion of patients who survive day i among those who survive day i-1. Therefore it can be estimated according to

$$\hat{P}_1 = \frac{(\text{Total number of patients}) - (\text{Total number of events during day 1})}{(\text{Total number of patients})}$$

$$\hat{P}_i = \frac{(\text{Number of Patients at risk day i}) - (\text{Total number of events during day i})}{(\text{Number of Patients at risk day i})}$$

K-M Estimate: General Formula

- Rank the survival times as $t_{(1)} \le t_{(2)} \le \ldots \le t_{(n)}$.
- Formula

$$\hat{S}(t) = \prod_{t_{(i)} \le t} \frac{n_i - d_i}{n_i}$$

- n_i patients at risk
- d_i failures

$$P_1 = \frac{19}{20} = 0.95 \qquad S(1) = S(0) \times P_1 = 1 \times 0.95 = 0.95$$

$$P_3 = \frac{17}{19} = 0.89474 \qquad S(3) = S(0) \times P_1 \times P_3 = 1 \times \frac{19}{20} \times \frac{17}{19} = 1 \times 0.95 \times 0.89464 = 0.85$$

In SAS: PROC LIFETEST

Using SAS

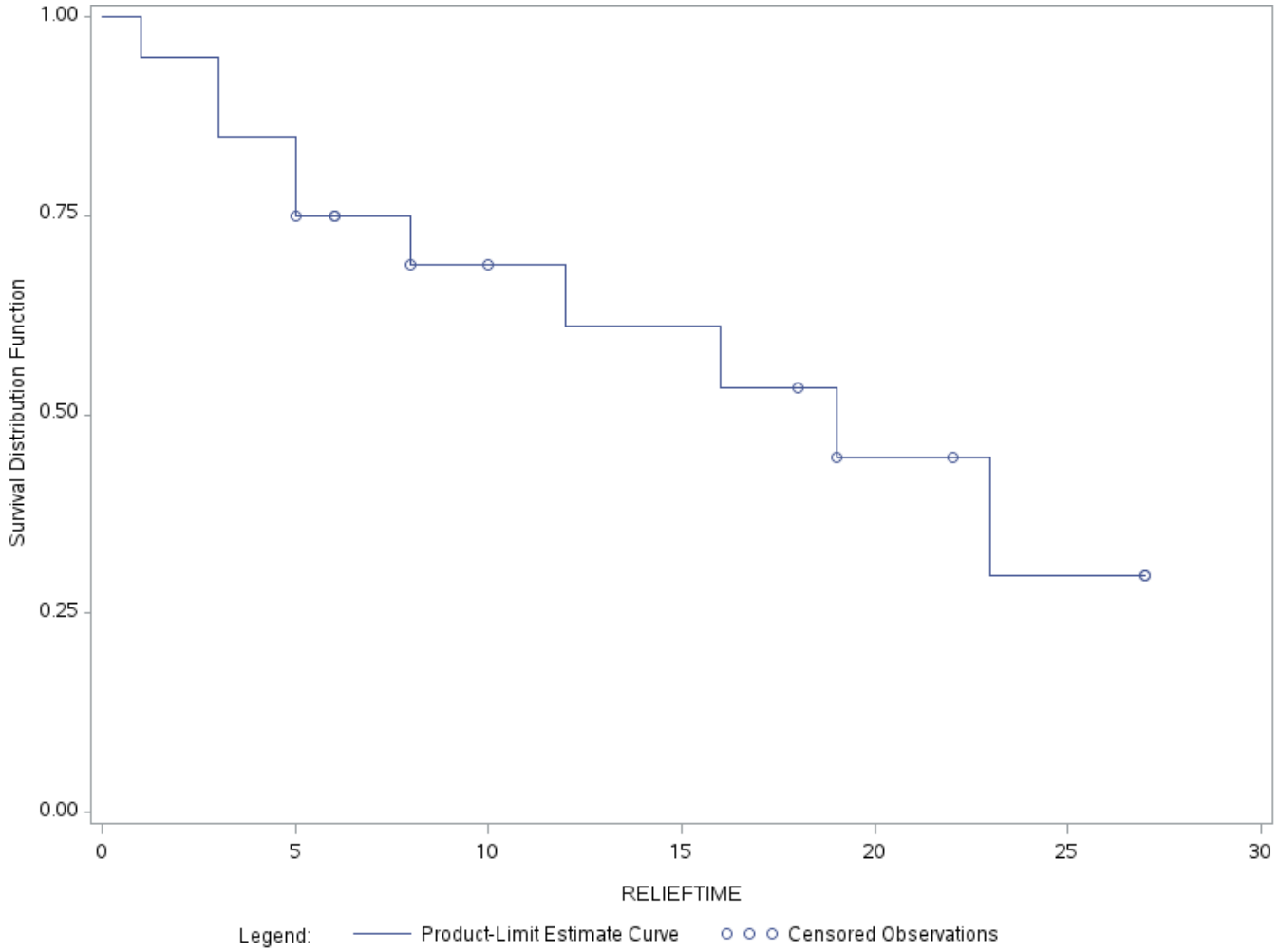
```
data survival;
input INDIVIDUAl STARTDAY LASTDAY RELIEFTIME STATUS CENS;
cards;
1 1 28 27 0 1
2 1 28 27 0 1
3 1 6 5 1 0
4 1 9 8 1 0
5 1 24 23 1 0
6 2 18 16 1 0
7 5 8 3 1 0
8 5 24 19 1 0
9 6 28 22 0 1
10 9 28 19 0 1
11 10 15 5 1 0
12 10 22 12 1 0
13 10 28 18 0 1
14 18 28 10 0 1
15 20 28 8 0 1
16 22 28 6 0 1
17 22 28 6 0 1
18 23 28 5 0 1
19 24 27 3 1 0
20 27 28 1 1 0
;
run;
```

```
proc lifetest data=survival
plots=(s);
TIME relieftime*cens(1);
run;
```

The LIFETEST Procedure

Product-Limit Survival Estimates						
RELIEFTIME		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		1.0000	0	0	0	20
1.0000		0.9500	0.0500	0.0487	1	19
3.0000		.	.	.	2	18
3.0000		0.8500	0.1500	0.0798	3	17
5.0000		.	.	.	4	16
5.0000		0.7500	0.2500	0.0968	5	15
5.0000	*	.	.	.	5	14
6.0000	*	.	.	.	5	13
6.0000	*	.	.	.	5	12
8.0000		0.6875	0.3125	0.1070	6	11
8.0000	*	.	.	.	6	10
10.0000	*	.	.	.	6	9
12.0000		0.6111	0.3889	0.1193	7	8
16.0000		0.5347	0.4653	0.1265	8	7
18.0000	*	.	.	.	8	6
19.0000		0.4456	0.5544	0.1332	9	5
19.0000	*	.	.	.	9	4
22.0000	*	.	.	.	9	3
23.0000		0.2971	0.7029	0.1503	10	2
27.0000	*	.	.	.	10	1
27.0000	*	.	.	.	10	0

The LIFETEST Procedure



Comparing Survival Functions

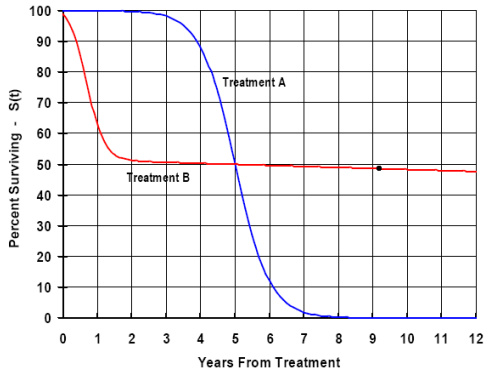
- Question: Did the treatment make a difference in the survival experience of the two groups?
- Hypothesis: H_0: $S_1(t)=S_2(t)$ for all $t \geq 0$.
- Two tests often used :
  1. Log-rank test (Mantel-Haenszel Test);
  2. Cox regression

TWO DIFFERENT SURVIVAL CURVES

These two survival curves have the same 5-year survival rate of 50%.

The interpretation of the curves is substantially different, however.

Which treatment is preferable will be a subjective judgement.



Using SAS

```
data survival;
input INDIVIDUAl TIME CENS GROUP $;
cards;
20 2 0 DRUG
7 4 0 DRUG
19 4 0 DRUG
3 6 0 CONTROL
11 6 0 DRUG
18 6 1 DRUG
15 7 1 CONTROL
17 7 1 CONTROL
4 9 0 DRUG
16 9 1 CONTROL
14 11 1 DRUG
12 13 0 CONTROL
6 17 0 DRUG
13 19 1 DRUG
8 20 0 DRUG
10 20 1 CONTROL
9 23 1 CONTROL
5 24 0 CONTROL
1 28 1 CONTROL
2 28 1 CONTROL
;
run;
```

```
proc lifetest data=survival
plots=(s);
TIME time*cens(1);
strata GROUP;
run;
```

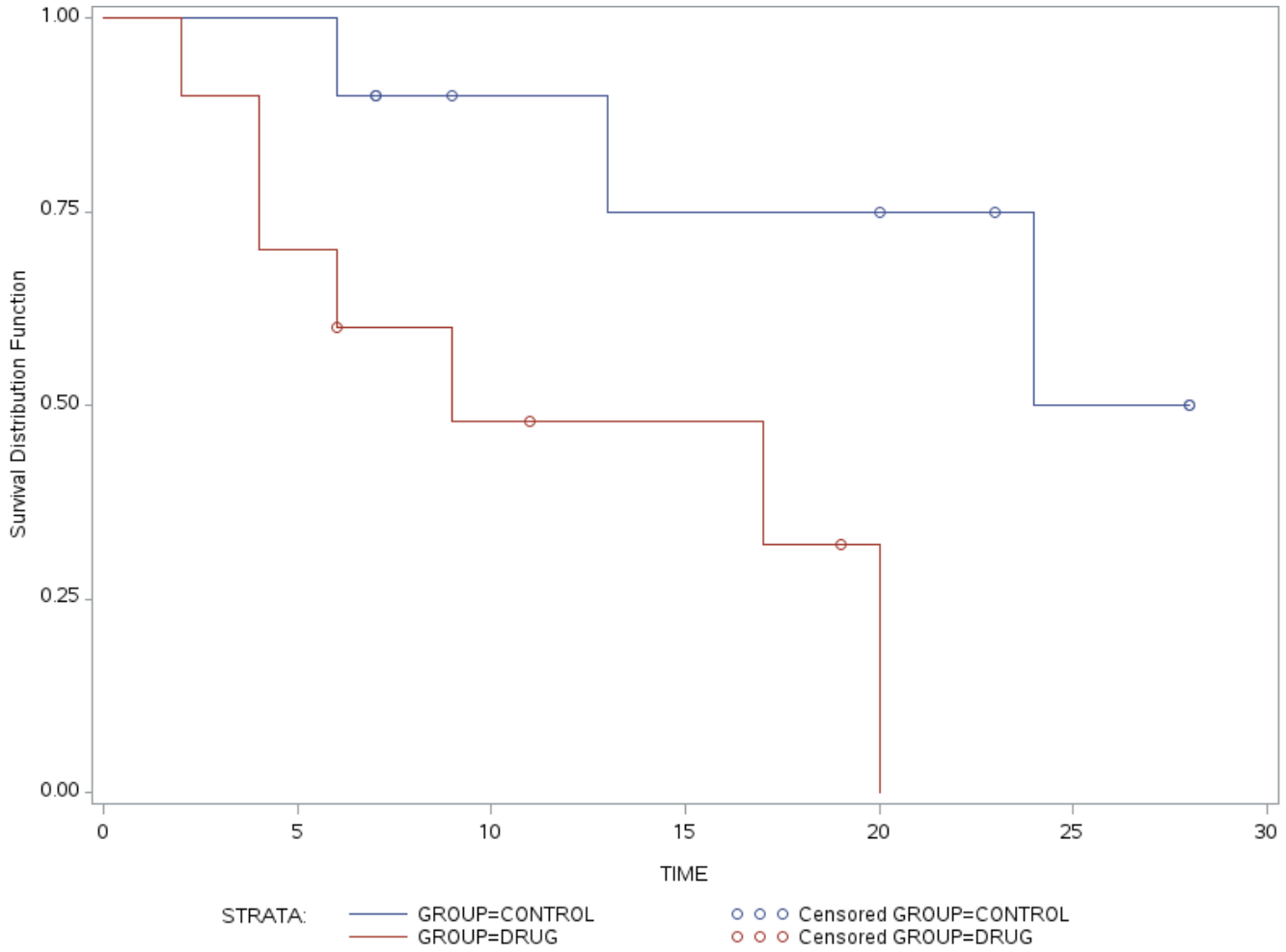
The LIFETEST Procedure
Stratum 1: GROUP = CONTROL

Product-Limit Survival Estimates						
TIME		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		1.0000	0	0	0	10
6.0000		0.9000	0.1000	0.0949	1	9
7.0000	*	.	.	.	1	8
7.0000	*	.	.	.	1	7
9.0000	*	.	.	.	1	6
13.0000		0.7500	0.2500	0.1581	2	5
20.0000	*	.	.	.	2	4
23.0000	*	.	.	.	2	3
24.0000		0.5000	0.5000	0.2297	3	2
28.0000	*	.	.	.	3	1
28.0000	*	.	.	.	3	0

The LIFETEST Procedure
Stratum 2: GROUP = DRUG

Product-Limit Survival Estimates						
TIME		Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000		1.0000	0	0	0	10
2.0000		0.9000	0.1000	0.0949	1	9
4.0000		.	.	.	2	8
4.0000		0.7000	0.3000	0.1449	3	7
6.0000		0.6000	0.4000	0.1549	4	6
6.0000	*	.	.	.	4	5
9.0000		0.4800	0.5200	0.1640	5	4
11.0000	*	.	.	.	5	3
17.0000		0.3200	0.6800	0.1703	6	2
19.0000	*	.	.	.	6	1
20.0000		0	1.0000	.	7	0

The LIFETEST Procedure



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.8681	1	0.0154
Wilcoxon	4.6579	1	0.0309
-2Log(LR)	4.4006	1	0.0359

Limitation of Kaplan-Meier curves

- What happens when you have several covariates that you believe contribute to time-to-event?
- Example
  - Smoking, hyperlipidemia, diabetes, hypertension, contribute to time to myocardial infarct
- Can use **stratified K-M curves** – but the combinatorial complexity of more than two or three covariates prevents practical use
- Need another approach – multivariate **Cox proportional hazards model** is most commonly used
  - (think multivariate regression or logistic regression)

Part II: Cox Regression

- Introduction to the proportional hazard model (PH)
- Comparing two groups
- A numerical example

Cox Regression

- In 1972 Cox suggested a model for survival data that would make it possible to take covariates into account. Up to then it was customary to discretise continuos variables and build subgroups.
- Cox idea was to model the hazard rate function

$$P(t \leq T < t + \Delta t \mid T \geq t) = h(t)\Delta t$$

where h(t) is to be understood as an intensity i.e. a probability by time unit. Multiplied by time we get a probability. Think of the analogy with speed as distance by time unit. Multiplied by time we get distance.

The model

$$h_i(t) = h_0(t)e^{\boldsymbol{\beta}^T \mathbf{z}_i}$$

$\boldsymbol{\beta}^T = (\beta_1, \quad \ldots \quad , \beta_k)$ the parameter vector;

$\mathbf{Z_i} = (Z_{i1}, \quad \ldots \quad , Z_{ik})$ the covariate vector.

where each parameter is a measure of the importance of the corresponding variable.

Two individuals with different covariate values will have hazard rate functions which differ by a multiplicative term. The hazards are propotional; therefore called *proportional hazard model*

Note that the hazard ratio is constant in t.

$$h_i(t) = h_0(t)e^{\beta^T z_i}, \quad i = 1,2$$

$$\frac{h_1(t)}{h_2(t)} = \frac{h_0(t)e^{\beta^T z_1}}{h_0(t)e^{\beta^T z_2}} = e^{\beta^T(z_1 - z_2)} = C$$

$$h_1(t) = C\, h_2(t);$$

Cox Regression Model

- Semiparametric model (due to the fact that h_0 is not explicitly modeled)
- No specific distributional assumptions (but includes several important parametric models as special cases).
- Can handle both continuous and categorical predictor variables (think: logistic, linear regression)
- Parameters are estimated based on partial likelihood (not full ML-estimation).

Cox proportional hazards model, continued

- Maximum partial likelihood estimates are not fully efficient, but share other general properties of ML-estimates
  - Asymptotic sampling varainces can be estimated
  - Likelihood ratio tests, Wald and score tests can be constructed for testing of the β-parameters
- The β-parameters can be interpreted in terms of **hazard ratio,** a relative risk measure
- Easy implementation (SAS procedure PHREG).

- Parametric approaches are an alternative, but they require stronger assumptions about h(t).

Example

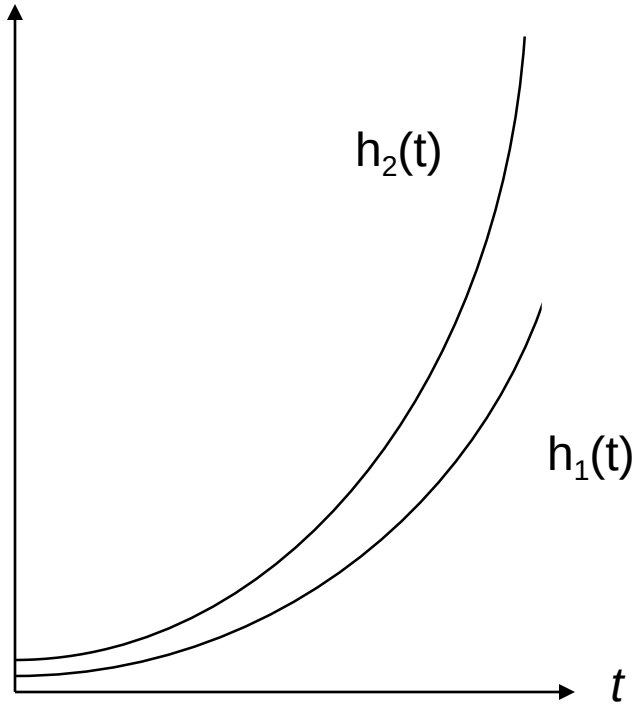
Assume we have a situation with one covariate that takes two different values, 0 and 1. This is the case when we wish to compare two treatments

$$k = 1; \boldsymbol{\beta} = \beta_1 = \beta;$$

$$\mathbf{z}_1 = \mathbf{0}; \mathbf{z}_2 = 1;$$

$$h_1(t) = h_0(t);$$

$$h_2(t) = h_0(t)e^{\beta}$$



A numerical Example

TIME TO RELIEF OF ITCH SYMPTOMS FOR PATIENTS USING A STANDARD AND EXPERIMENTAL CREAM

PATIENT #	DRUG	START DATE	STOP DATE	RELIEF TIME	STATUS
1	1	1	16	15	1
2	1	1	18	17	1
3	2	1	20	19	1
4	1	4	18	14	1
5	1	4	20	16	1
6	1	4	22	18	1
7	2	4	24	20	1
8	1	8	25	17	1
9	1	8	29	21	1
10	2	8	28	20	1
11	2	8	30	22	1
12	1	11	30	19	1
13	2	11	30	19	1
14	2	11	31	20	0
15	1	15	31	16	0
16	2	15	31	16	0
17	2	15	31	16	0
18	1	18	31	13	0
19	2	18	31	13	0
20	2	18	31	13	0

Using SAS

```
data survival;
input PATIENT DRUG START STOP RELIEF STATUS;
cards;
1 1 1 16 15 1
2 1 1 18 17 1
3 2 1 20 19 1
4 1 4 18 14 1
5 1 4 20 16 1
6 1 4 22 18 1
7 2 4 24 20 1
8 1 8 25 17 1
9 1 8 29 21 1
10 2 8 38 20 1
11 2 8 30 22 1
12 1 11 30 19 1
13 2 11 30 19 1
14 2 11 31 20 0
15 1 15 31 16 0
1 2 15 31 16 0
17 2 15 31 16 0
18 1 18 31 13 0
19 2 18 31 13 0
20 2 18 31 13 0
;
run;
```

```
PROC TTEST DATA=SURVIVAL;
CLASS DRUG;
VAR RELIEF;
RUN;

PROC PHREG DATA=SURVIVAL;
MODEL RELIEF * STATUS(0) = DRUG;
run;
```

TTEST OF EQUALITY OF MEANS

DRUG	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
1		10	16.6000	2.3664	0.7483	13.0000	21.0000
2		10	17.8000	3.1198	0.9866	13.0000	22.0000
Diff (1-2)	Pooled		-1.2000	2.7689	1.2383		
Diff (1-2)	Satterthwaite		-1.2000		1.2383		

DRUG	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
1		16.6000	14.9072	18.2928	2.3664	1.6277	4.3202
2		17.8000	15.5682	20.0318	3.1198	2.1459	5.6956
Diff (1-2)	Pooled	-1.2000	-3.8015	1.4015	2.7689	2.0922	4.0947
Diff (1-2)	Satterthwaite	-1.2000	-3.8151	1.4151			

The mean difference in the time to “cure” of 1.2 days is not statistically significant between the two groups.

Summary of the Number of Event and Censored Values			
Total	Event	Censored	Percent Censored
20	13	7	35.00

Model Fit Statistics		
Criterion	Without Covariates	With Covariates
-2 LOG L	50.914	46.111
AIC	50.914	48.111
SBC	50.914	48.676

Testing Global Null Hypothesis: BETA=0			
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4.8032	1	0.0284
Score	5.0192	1	0.0251
Wald	4.4647	1	0.0346

Analysis of Maximum Likelihood Estimates						
Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
DRUG	1	-1.33960	0.63398	4.4647	0.0346	0.262

Interpretation

- The estimate of the DRUG variable is -1.3396 with a p-value of 0.0346.
- The negative sign indicates a negative association between the hazard of being cured and the DRUG variable.
- DRUG is coded 1 for the *new drug* and coded 2 for the *standard drug*. Therefore the hazard of being cured is lower in the group given the standard drug.
- This is an awkward but accurate way of saying *that the new drug tends to produce a cure more quickly than the standard drug*.
- The mean time to cure is lower in the group given the new drug. There is an inverse relationship between the average time to an event and the hazard of that event.

The ratio of the hazards is given by

$$\frac{h_2(t)}{h_1(t)} = \frac{e^{\hat{\beta}\times 2}}{e^{\hat{\beta}\times 1}} = e^{(2-1)\hat{\beta}} = e^{-1.3396} = 0.262$$

- At each time point the cure rate of the standard drug is about 25% of that of the new drug. Put more positively, we might state that the cure rate is 3.8 times better in the group given the experimental cream compared to the group given the standard cream.

Generalizations of Cox regression

1. Time dependent covariates
2. Stratification
3. General link function
4. Likelihood ratio tests
5. Sample size determination
6. Goodness of fit
7. SAS

References

- Cox & Oakes (1984) “Analysis of survival data”. Chapman & Hall.
- Fleming & Harrington (1991) “Counting processes and survival analysis”. Wiley & Sons.
- Allison (1995). “Survival analysis using the SAS System”. SAS Institute.
- Therneau & Grambsch (2000) “Modeling Survival Data”. Springer.
- Hougaard (2000) “Analysis of Multivariate survival data”. Springer.