ST227 Survival Models - Part IV
Estimating the Lifetime Distribution

Department of Statistics,
London School of Economics

March 3, 2019

1 The Kaplan-Meier Estimate and the Cox Regression Model

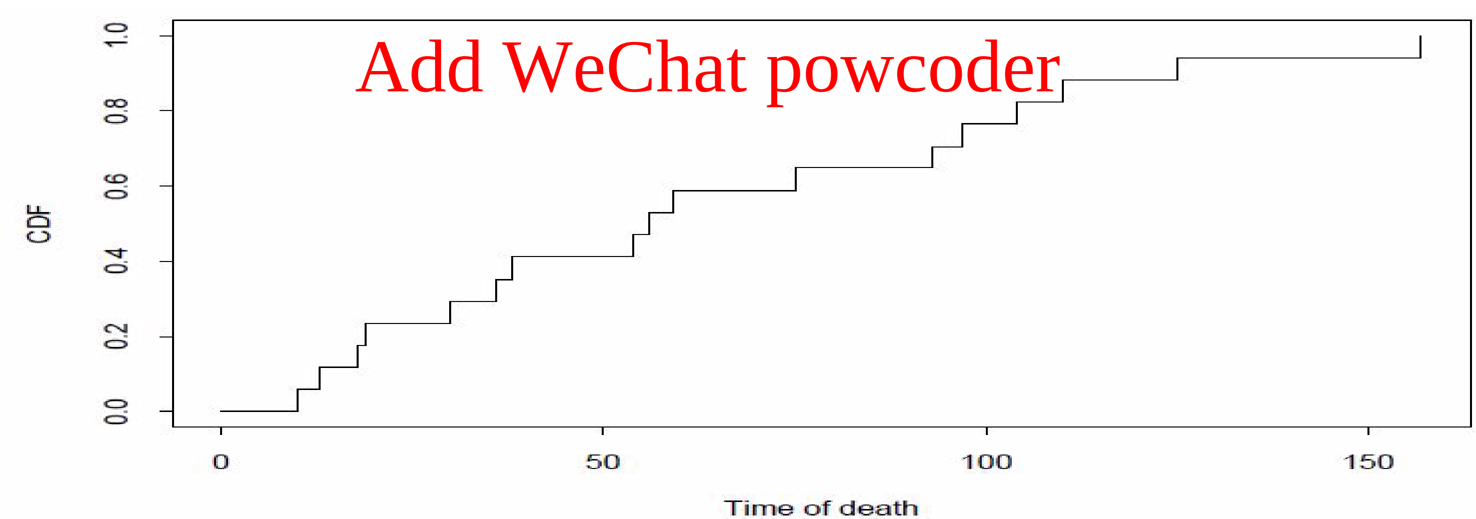
1.1 Cohort Studies

- We want to estimate the distribution of the lifetime variables T, T_0 and T_x. We consider first data that arises from a **cohort study**. Later we will look at the **cross-sectional study**.

- We observe a **large number of new-born lives**. The **proportion alive** at age t is an estimate of the survivor function $S(t)$. This will be a step function, and the larger the number of lives the smoother the estimate of $S(t)$ will be.

- **Equivalently**, we can take the **proportion who have died** up to and including time t. This estimate of $F(t)$ is the **empirical distribution function**, $\hat{F}(t)$. A typical graph of $\hat{F}(t)$(for a small sample) would look like this.

Figure 1. Empirical distribution function, $\hat{F}(t)$



- This graph could be smoothed or **graduated**, if required.

- **Problems**: the cohort study suffers from a number of problems which make it unsuitable in insurance.

  - To observe the mortality of a cohort over all ages requires around **100 years of data**.

  - In practice, we will lose track of a large number of individuals in the cohort. This may produce a biased sample, since the mortality of the lost lives may be different from those lives that remain in the study. We call this problem **censoring**. For censored lives, all we know is that these lives had lifetimes that exceed a certain age.

1.2 Censoring

- As was mentioned in Part III, censoring is a key feature of survival data and has a profound impact on the estimation procedure.

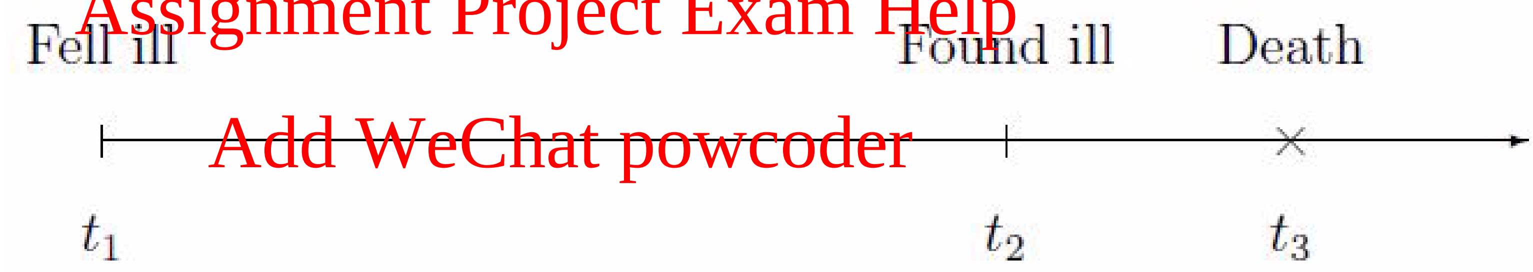
- In particular, it is a form of missing data problem which arises when we do not observe the exact length of a lifetime, but observe only that its length falls within some interval.

- This can happen in several ways.

1. **Right-censoring** occurs if the censoring mechanism cuts short observations in progress.

Example: An example of this is the ending of a mortality investigation before all the lives being observed have died. Persons still alive when the investigation ends are right censored. We know only that their **lifetimes exceed the censoring time** t, i.e. $T \geq t$.

2. **Left-censoring** occurs if we do not know the time of entry into the state of interest.

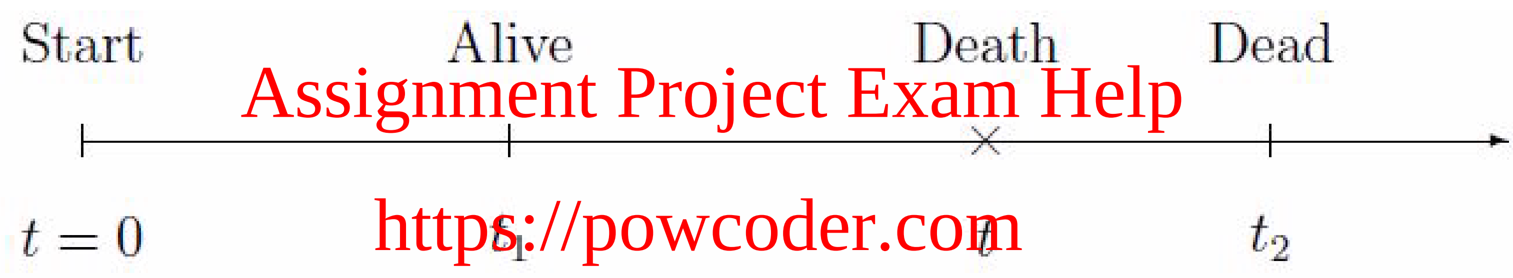


Example: Left-censoring often in medical studies in which patients are subject to regular examinations. Discovery of a condition tells us only that the onset fell in the period since the previous examination; the time elapsed since onset has been left-censored. This time we can say that the survival time from t_1, the **unknown time of onset**, exceeds $t_3 - t_2$, i.e. $T \geq t_3 - t_2$.

Assignment Project Exam Help
https://powcoder.com
Add WeChat powcoder

3. **Interval-censoring** occurs when the event of interest occurs somewhere between two times t_1 and t_2. An example arises in mortality investigations, where we might know only the calendar year of death. In the diagram we

know the life is alive at time t_1 and dead at time t_2, but we do not know the exact time of death t. All we can say is that $t_1 \leq T \leq t_2$.



Note: Both right- and left-censoring can be seen as special cases of interval censoring.

Further terminology: There are a number of ways in which any of the above forms of censoring can arise.

1. **Random censoring** occurs if the censoring time C_i for life i is a random variable. The observation is then censored if $C_i < T_i$, where T_i is the random lifetime of life i.

   Example: In medical studies a patient moves away from the study area at a random time, and the study loses track of the patient.

2. **Non-informative censoring** occurs if the censoring gives no information about the lifetimes $\{T_i\}$. Thus, if each member of the pair $\{T_i, C_i\}$ is independent then censoring is non-informative.

   Example: In medical studies a patient moves away from the study area at a time unconnected with the progress of their illness. This is non-informative. However, if the patient moves away from the study area in order to receive a different treatment this is informative. Informative censoring may introduce bias into the observed survival times and must be treated with extra care.

3. **Type I** censoring occurs when the censoring times are known in advance, i.e. censoring occurs at fixed times.

   Example: A mortality investigation may end on a particular date.

4. **Type II** censoring occurs if observation continues until a fixed known number of deaths occurs. In this case the number of events is non-random.

   Example: Type II censoring occurs in reliability testing when components may be tested until a certain number fail. This kind of censoring is uncommon in mortality studies.

1.3 Cross-sectional Studies

Instead of following a **cohort** of lives for a long time we can use a **cross-sectional** study. We divide the investigation into single years of age, say, $x \to x + 1$, and then follow each mini "cohort" for three or four years. There are a number of points to note:

- We are now mixing mortality since the mortality of a 60 year old now is not the same as the mortality of a 57 year old in three years time.

- The method introduces type I censoring in the most obvious way, since most lives will not die (we suppose).

- In the following section we develop the empirical distribution function to allow for censoring.

- We will consider lifetimes as a function of time t without mention of a starting age x.

- The following could be applied equally to newborn lives, to lives aged x at outset, or to lives with some property in common at time $t = 0$, for example diagnosis of a medical condition.

- Medical studies are often based on time since diagnosis, or time since the start of treatment, and if the patient's age enters the analysis it is usually as an explanatory variable in a regression model.

1.4 The Kaplan-Meier (K-M) estimate of the survivor function

Suppose we have n lives with **survival times** $t_1, ..., t_n$ where some of the **observations are right-censored (non-informative)**. Suppose there are r distinct **death times**, $r \le n$, which we arrange in ascending order $t_{(1)} < t_{(2)} < ... < t_{(r)}$. We form the set of intervals

$$I_j = [t_{(j)}, t_{(j+1)}), \qquad j = 0, 1, ..., r$$

where $t_{(0)} = 0$ and $t_{(r+1)} = \infty$. Let

$$n_j = \text{number of individuals alive at } t_{(j)}-$$

i.e., immediately before time $t_{(j)}$ (the jth death-time), and let

$$d_j\text{=number of deaths at } t_{(j)}$$

- Consider the interval $(t_{(j)} - \delta, t_{(j)}]$. There are n_j patients **alive at** $t_{(j)} - \delta$, i.e. the beginning of the interval, and d_j **deaths at** $t_{(j)}$.

- Thus, the **probability that a patient dies in** $(t_{(j)} - \delta, t_{(j)}]$ is estimated by d_j/n_j .

- So the maximum likelihood estimate of the **probability of surviving** from $t_{(j)} - \delta$ to $t_{(j)}$ is $1 - d_j/n_j = (n_j - d_j)/n_j$.

- Further, since there are **no deaths in** $(t_{(j)}, t_{(j+1)} - \delta]$, the **probability of surviving** from $t_{(j)} - \delta$ to $t_{(j+1)} - \delta$ is also $(n_j - d_j)/n_j$. If we let $\delta \longrightarrow 0$ then we can estimate the probability of surviving the interval $t_{(j)}$ to $t_{(j+1)}$ as $(n_j - d_j)/n_j$.

We now assume that all deaths are independent of each other. Let time t lie in I_k. Then we estimate the probability of survival to time t as the probability of surviving through $t_{(k)}$ to $t_{(k+1)}$, and all preceding intervals. This gives the maximum likelihood estimate of the survivor function, usually known as the **Kaplan-Meier estimate**,

$$\widehat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j}\right), t \in [t_{(k)}, t_{(k+1)}). \tag{1}$$

with $\widehat{S}(0) = 1$. Note the following

- If the largest observation corresponds to a death then $n_r = d_r$ so $\widehat{S}(t) = 0$ for $t \geq t_r$.

- If the largest observation, say t^*, is censored then $\widehat{S}(t)$ is undefined for $t \geq t^*$.

Example: The data refer to an experiment to study the use of an IUD device. The time origin corresponds to the first day on which the woman uses the IUD. The observation is the time to discontinuation of use of the IUD because of bleeding problems. Notice that some observations are *censored*, denoted * . In this example, censoring could occur for a number of reasons: (a) desire for pregnancy, (b) no further need of contraception, (c) lost to follow-up, (d) the study ended.

Table 1. Time in weeks to discontinuation of the use of an IUD

10	13*	18*	19	23*	30	36	38*	54*
56*	59	75	93	97	104*	107	107*	107*

We set out the calculation of the Kaplan-Meier estimate as follows.

Table 2. Kaplan-Meier estimate of the survivor function

Interval	n_j	d_j	$(n_j - d_j)/n_j$	$\widehat{S}(t)$
0-	18	0	1.0000	1.0000
10-	18	1	0.9444	0.9444
19-	15	1	0.9333	0.8815
30-	13	1	0.9231	0.8137
36-	12	1	0.9167	0.7459
59-	8	1	0.8750	0.6526
75-	7	1	0.8571	0.5594
93-	6	1	0.8333	0.4662
97-	5	1	0.8000	0.3729
107	3	1	0.6667	0.2486

The table is constructed in the following steps:

1. **Order the data**. If there are any **censored observations** with the same times as a death time **place them after the deaths** d_j. Here d_j is the number of individuals experiencing the event at duration **distinct death times**, $t_{(j)}$, $j = 1, ..., r$.

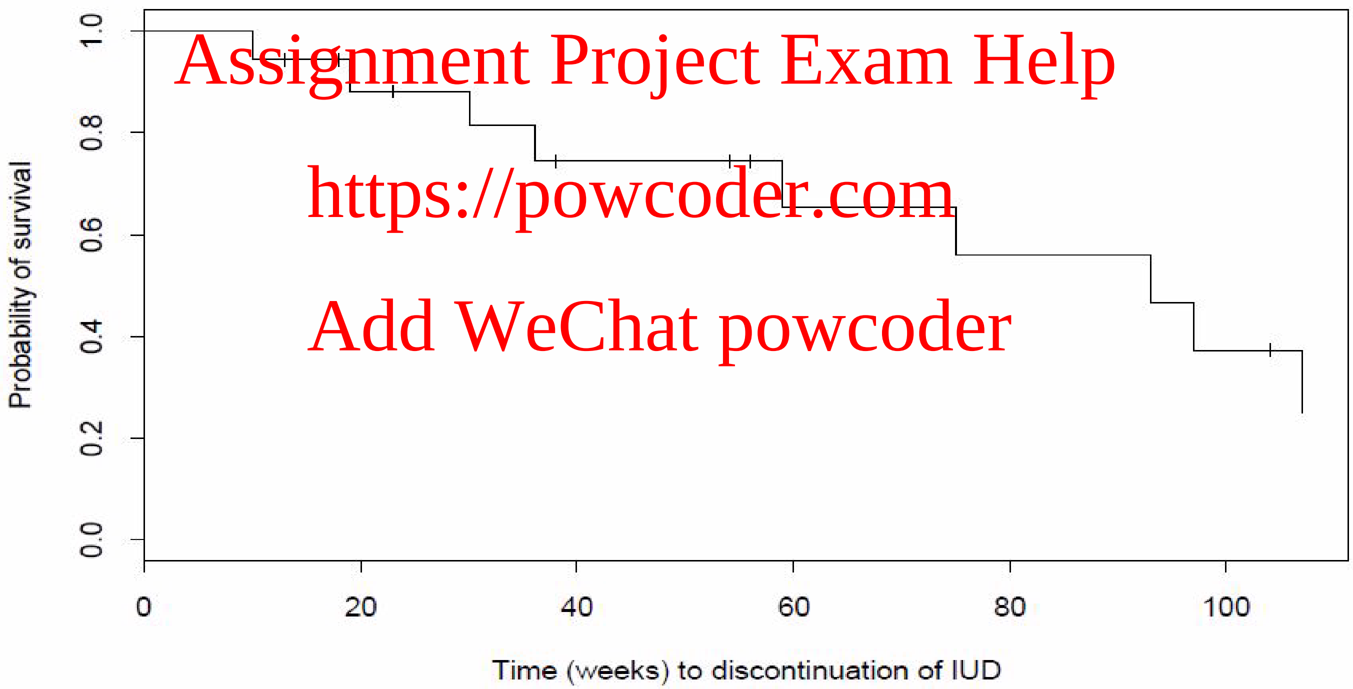
2. Column one consists of the distinct death times, $t_{(j)}$, $j = 1, ..., r$.

3. Column two consists of the number of observations n_j greater than or equal to the death time $t_{(j)}$ (including the deaths at time $t_{(j)}$).

4. Column three consists of the number of deaths at time $t_{(j)}$; if there are no multiple deaths the entries will all be 1 (except the first which is 0).

5. Columns four and five compute $\widehat{S}(t)$.

Figure 2. Kaplan-Meier estimate of survivor function, $\widehat{S}(t)$, for IUD data



The standard error of the Kaplan-Meier estimate

How good is the Kaplan-Meier estimate? We can answer this by computing its standard error. We could use this to compare two lifetime distributions, for example. The variance is given by Greenwood's formula (proof not required)

$$Var(\widehat{S}(t)) \approx (\widehat{S}(t))^2 \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}, \quad t \in [t_{(k)}, t_{(k+1)}). \tag{2}$$

1.5 The Cox Regression Model

- The non-parametric approach of the Kaplan-Meier (K-M) estimate is not well suited to answering questions concerning the **effects of covariates** on the survival function.

- A **covariate** is any **quantity recorded in respect of each life**, such as age, sex, type of treatment, level of medication, severity of symptoms and so on.

- We will assume that the covariates in respect of the ith life are represented by a $1 \times p$ vector $\left(z_{i1}, ..., z_{ip}\right)$.

- If a covariate partitions the lives into a small number of homogeneous groups, eg: Sex, then we may find a K-M estimate for each group.

- But, **if the covariate assumes many values**, eg: Age, then this approach is not possible. We want a **regression type model**.

- Assume that $(\beta_1, \beta_2, ..., \beta_p)$ is the **unknown vector of p regression coefficients** corresponding to the covariates.

- In this section we consider the **proportional hazards model** or the **Cox regression model** (after D. R. Cox who invented it (1972)).

Definition: A survival time T follows a **Cox proportional hazards** model if the **hazard function (force of mortality)** for the ith life with covariate $z_i = (z_{i1}, z_{i2}, ..., z_{ip})'$ can be written as

$$\lambda(t; z_i) = \lambda_0(t) \exp(\beta' z_i) = \lambda_0(t) \exp(\beta_1 z_{i1} + \beta_2 z_{i2+...+} \beta_p z_{ip}). \qquad (3)$$

Comments

1. The function $\lambda_0(t)$ is known as the **baseline hazard,** the hazard for an individual with a covariate vector equal to zero.

   This means that we set $\beta_0 = 0$, since any fixed effect can simply be absorbed into the baseline hazard.

2. The covariates act **multiplicatively** on the baseline hazard; equivalently they act **additively** on the log scale.

3. **Cox** showed that it is possible to **estimate the effects of the covariates without specifying the baseline hazard** $\lambda_0(t)$.

- Under the **Cox model**, the hazards of **different lives with covariate vectors** z_1 and z_2 are in the same proportion at all times t:

$$\frac{\lambda(t; z_1)}{\lambda(t; z_2)} = \frac{\exp(\beta' z_1)}{\exp(\beta' z_2)}, \tag{4}$$

where $\beta' = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_p \end{pmatrix}$, also (4) gives the **Hazard Ratio** and the **life in the denominator** is the **reference life**.

- The utility of this model arises from the fact that the general ‘shape’ of the hazard function for all individuals is determined by the baseline hazard, while the exponential term accounts for differences between individuals.

- Thus, **if we know the** β's **we can compare the lives**, since the β's tell us which lives have high/low hazards, and this **without knowing the hazard function** $\lambda_0(t)$!

- So, how do we estimate β? : The **partial likelihood** estimates the regression coefficients but avoids the need to estimate $\lambda_0(t)$.

The partial likelihood function, $L(\beta)$

- First, we set up some **notation**.

- We suppose that data are available on $\mathbf{n}$ **lives**, amongst whom there are $\mathbf{r}$ distinct **death (or failure) times** and $\mathbf{n}-\mathbf{r}$ **right-censored survival**

times. We will assume there are *no ties in the death times*; ties lead to complications so we will avoid this point.

- We denote the r ordered death times by $t_{(1)} < t_{(2)} < ... < t_{(r)}$.

- The set of lives who are at risk at time t_j is denoted $R(t_{(j)})$, so that $R(t_{(j)})$ is the **set of lives who are alive and uncensored** just prior to $t_{(j)}$. We call $R(t_{(j)})$ the **risk set** at time $t_{(j)}$. Suppose the life with explanatory variable z_j dies at time t_j.

Result: The (partial) likelihood function for the Cox proportional hazards model is

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' z_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' z_l)} \tag{5}$$

Comments

1. Maximisation of this expression has to proceed numerically, and we will see how we can use the **survival package in R** for fitting a Cox model.

2. The partial likelihood behaves very like an ordinary likelihood function it furnishes all the statistical information needed for standard inference on the regression coefficients. Suppose $\widehat{\beta}$ is the estimate of β found by maximising $L(\beta)$, or equivalently $l(\beta)$, the log partial likelihood, and β_0

is the true unknown value of β. Then, viewed as a random variable, we have asymptotically

- $E\left(\widehat{\beta}\right) = \beta_0$;

- $Var\left(\widehat{\beta}\right) = -1/E(\frac{\partial^2 l}{\partial \beta^2})]_{\beta=\beta_0} \approx -1/\frac{\partial^2 l}{\partial \beta^2}]_{\beta=\widehat{\beta}}$.

3. The derivation of $L(\beta)$ makes no direct use of the actual censored and uncensored survival times. The censored and uncensored survival times only enter into the summation over the risk sets at the death times. For this reason $L(\beta)$ is not a true likelihood and it is usually referred to as a *partial likelihood*.

4. The likelihood function depends only on the ranking of the death and censoring times, since this determines the risk set at each death time.

Example 1: Suppose the lifetime, the age and the sex of a number of lives are recorded and the Cox model is fitted by maximising the partial likelihood. The fitted parameters are:$\beta_f = -0.5$ where f denotes a female life and $\beta_\alpha = 0.01$ is the effect of age in years. Taking a **male aged 40 as the reference life**, find the ratio of the hazards for (a) a male aged 20 (b) a male aged 60 (c) a female aged 20 (d) a female aged 60.

- In this case we have two covariates **sex** which is **categorical** and we assume that it takes the value "0" for males and the value "1" for females **age** which is **continuous** and takes the values, 40 (denominator), 20 and 60 for a male in (a) and (b) and 20 and 60 for a male in (c) and (d).

- Also, a male aged 40 is the **reference life** so it will go to the **denominator of the ratio of the hazards.**

- We want to compute the hazard ratio:

$$\frac{\exp(\beta' z_1)}{\exp(\beta' z_2)}$$

where $\beta' = (-0.5, 0.01)$, $z' = (0, age)$ for a male, and $z' = (1, age)$ for a female, where we consider that $z' \equiv z_1$ in the numerator, (cases (a), (b) and (c)) and $z' \equiv z_2$ in the denominator (reference life.)

- Note that the **denominator** in each case will be $\exp(\beta' z_2) =$

$$= \exp(-0.5 \times 0 + 40 \times 0.01) = \exp(40 \times 0.01) = 1.491825.$$

- We have

(a) $\exp(-0.5 \times 0 + 20 \times 0.01)/\exp(40 \times 0.01) = \exp(-0.2) = 0.82$

(b) $\exp(-0.5 \times 0 + 60 \times 0.01)/\exp(40 \times 0.01) = \exp(0.2) = 1.22$

(c) $\exp(-0.5 \times 1 + 20 \times 0.01)/\exp(40 \times 0.01) = \exp(-0.7) = 0.50$

(d) $\exp(-0.5 \times 1 + 60 \times 0.01)/\exp(40 \times 0.01) = \exp(-0.3) = 0.74$

Before giving our next example we give a reminder of the definition of the score function and define the **observed information function**.

Definition: We define the **score function** $U(\beta)$ by

$$U(\beta) = \frac{\partial l}{\partial \beta} \tag{6}$$

where $l = l(\beta) = Log(L(\beta))$ is the log-likelihood function.

Definition: We define the **observed information function** $Inf(\beta)$ by

$$Inf(\beta) = -\frac{\partial l^2}{\partial \beta^2} \tag{7}$$

Reminder: $I(\beta) = -E\left(\frac{\partial^2 l}{\partial \beta^2}\right)$ is the Fisher information function.

Example 2: The table gives the data on times to claim for permanent health insurance (PHI) policies for two groups of lives.

Male	2	5+	9	11+	14
Female	4	7+	10+	12+	15

where the + indicates that the observation was censored.

Solution

- We want to compute the partial likelihood in the Cox proportional hazards model which is given by

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta z_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta z_l)}$$

where, as was previously mentioned, $R(t_{(j)})$ is the **set of lives who are alive and uncensored** just prior to the r ordered **death (or failure) times** $t_{(1)} < t_{(2)} < ... < t_{(r)}$ and where $z^{'}$ (i.e. $z_{(j)}$ in the numerator and z_l in the denominator) is the vector of covariate information and β are the regression coefficients.

- Here $z' = sex$ and we assume $z' = 0$ for a male, and $z' = 1$ for a female. This means that each individual's multiplier (in the numerator and the denominator) is given by

$$e^{\beta z} = \begin{cases} 1, & \text{if } z = 0, \text{ i.e. males} \\ e^{\beta}, & \text{if } z = 1, \text{ i.e. females} \end{cases}$$

Step by Step approach for calculating $L(\beta)$

We have data on 10 individuals, 5 males and 5 females. Once more

Male	2	5+	9	11+	14
Female	4	7+	10+	12+	15

1. **Order the non-censored death times $t_{(j)}$.** Here, there are **5 non-censored (i.e. without a "+" sing) death times** in total, i.e. $j = 1, .., 5$. Of those 5 deaths, $d_1 = 3$ come from the **first group (males)** and $d_2 = 2$ come from the **second group (females)**.

   Hence, we get $t_{(1)} = 2 < t_{(2)} = 4 < t_{(3)} = 9 < t_{(4)} = 14 < t_{(5)} = 15$.

2. Next **we focus on** $R(t_{(j)})$. Let $r_j^{(1)}$ be the **number of lives from the first group (males)** who are **at risk just before the death times** $t_{(j)}$ and $r_j^{(2)}$ be the **number of lives from the second group (females)** who are **at risk just before the death times** $t_{(j)}$, $j = 1, .., 5$.

   - To **calculate** $r_j^{(1)}$ and $r_j^{(2)}$ we need to find **how many values form the Table above are greater than or equal to** $t_{(j)}$, $j = 1, .., 5$.

Once more, the Table is

Male	2	5+	9	11+	14
Female	4	7+	10+	12+	15

- The values of j, $t_{(j)}$, $r_j^{(1)}$ and $r_j^{(2)}$ are provided in a new Table below:

j	$t_{(j)}$	$r_j^{(1)}$	$r_j^{(2)}$
1	2	5	5
2	4	4	5
3	9	3	3
4	14	1	1
5	15	0	1

3. The partial likelihood is given by

$$L(\beta) = \prod_{j=1}^{5} \frac{\exp(\beta z_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta z_l)} = \frac{\left(e^{\beta \times 0}\right)^{d_1} \left(e^{\beta \times 1}\right)^{d_2}}{\prod_{j=1}^{5} \left(r_j^{(1)} e^{\beta \times 0} + r_j^{(2)} e^{\beta \times 1}\right)}$$

$$= \frac{\left(e^{\beta \times 0}\right)^{3} \left(e^{\beta \times 1}\right)^{2}}{\prod_{j=1}^{5} \left(r_j^{(1)} e^{\beta \times 0} + r_j^{(2)} e^{\beta' \times 1}\right)} = \frac{e^{2\beta}}{\prod_{j=1}^{5} \left(r_j^{(1)} + r_j^{(2)} e^{\beta}\right)},$$

$z' = 0$ and $d_1 = 3$ for a male, and $z' = 1$ and $d_2 = 2$ for a female.

Hence using the values of $r_j^{(1)}$ and $r_j^{(2)}$ from the above Table, for $j = 1, .., 5,$ we get

$$
\begin{aligned}
L(\beta) &= \frac{e^{2\beta}}{\prod_{j=1}^{5}\left(r_j^{(1)} + r_j^{(2)} e^{\beta}\right)} \\
&= \frac{e^{2\beta}}{\left(5+5e^{\beta}\right)\times\left(4+5e^{\beta}\right)\times\left(3+3e^{\beta}\right)\times\left(1+e^{\beta}\right)\times\left(0+e^{\beta}\right)} \\
&= \frac{e^{\beta}}{\left(5+5e^{\beta}\right)\times\left(4+5e^{\beta}\right)\times\left(3+3e^{\beta}\right)\times\left(1+e^{\beta}\right)\times\left(0+e^{\beta}\right)} \times e^{\beta} \\
&= e^{\beta}\times\frac{1}{5}\times\left(\frac{1}{1+e^{\beta}}\right)\times\frac{1}{4+5e^{\beta}}\times\frac{1}{3}\times\left(\frac{1}{1+e^{\beta}}\right)\times\frac{1}{1+e^{\beta}}\times 1 \\
&\propto \frac{e^{\beta}}{(1+e^{\beta})^3(4+5e^{\beta})} \ \text{(keep all the terms which involve } \beta\text{)} \\
&\Rightarrow l(\beta) = \beta - 3\log(1+e^{\beta}) - \log(4+5e^{\beta}) \ \text{(log-likelihood)}
\end{aligned}
$$

Estimation: The partial likelihood estimate of β satisfies

$$U(\beta) = \frac{\partial l}{\partial \beta} = 1 - \frac{3e^{\beta}}{1+e^{\beta}} - \frac{5e^{\beta}}{4+5e^{\beta}} = 0. \tag{8}$$

Trick 1: Putting $x = e^{\widehat{\beta}}$ we find $15x^2 + 8x - 4 = 0$ so $x = (\sqrt{76} - 4)/15 = 0.3145$ (since $x > 0$) and hence $\widehat{\beta} = -1.1567$.

What does $\widehat{\beta}$ mean? We have

$$\frac{\lambda_f(t)}{\lambda_m(t)} = \frac{\lambda_0(t)e^{\beta\times 1}}{\lambda_0(t)e^{\beta\times 0}} = \frac{\lambda_0(t)e^{\beta}}{\lambda_0(t)} = e^{\beta} = e^{-1.1567} = 0.3145 \tag{9}$$

and so $\lambda_f(t) \approx 0.3\lambda_m(t)$ for all t.

Standard error of $\widehat{\beta}$: We use

$$
\begin{aligned}
Var(\widehat{\beta}) &\approx 1/I(\widehat{\beta}) = -1/\left(\frac{\partial^2 l}{\partial \beta^2}\right)\Big]_{\beta=\widehat{\beta}} \\
&= -1/\left(\frac{\partial U}{\partial \beta}\right)\Big]_{\beta=\widehat{\beta}} \qquad (10)
\end{aligned}
$$

Trick 2: It is convenient to use the **chain rule** with $\upsilon = e^{\beta}$, to evaluate the expression $-\frac{\partial^2 l}{\partial \beta^2} = -\left(\frac{\partial U}{\partial \beta}\right)$ first.

Reminder: In calculus, the **chain rule** is a formula for computing the derivative of the composition $f \circ g = f\left(g\left(x\right)\right)$,of two functions say f and g.

It holds that

$$(f \circ g)' = f' \circ g \times g' = f'(g(x)) \times g'(x)$$

In our case

$$\begin{aligned}
\frac{d\upsilon}{d\beta} &= \frac{de^{\beta}}{d\beta} = e^{\beta}\frac{d\beta}{d\beta} = e^{\beta} = \upsilon \\
\implies & \frac{d\upsilon}{\upsilon} = d\beta \\
\implies & \frac{dU(\beta)}{d\beta} = \frac{\partial^2 l}{\partial \beta^2} = \frac{dU(\beta)}{d\upsilon/\upsilon} = \upsilon\frac{dU(\beta)}{d\upsilon}, \text{ where from (15):}
\end{aligned}$$

$$
\begin{aligned}
U(\beta) &= \frac{\partial l}{\partial \beta} = 1 - \frac{3e^{\beta}}{1+e^{\beta}} - \frac{5e^{\beta}}{4+5e^{\beta}}. \\
&= 1 - \frac{3\upsilon}{1+\upsilon} - \frac{5\upsilon}{4+5\upsilon}
\end{aligned}
$$

Hence to calculate $-\left(\frac{\partial^2 l}{\partial \beta^2}\right)$ we need to replace e^{β} with υ in (15 or above), then take the derivative of all terms w.r.t υ and multiply this by υ, we get:

$$
\begin{aligned}
-\frac{\partial^2 l}{\partial \beta^2} &= -\left(\frac{\partial U}{\partial \beta}\right) = \left[\frac{d}{dv}\left(-1 + \frac{3v}{1+v} + \frac{5v}{4+5v}\right)\right] v \\
&= \left[\frac{d}{dv}\left(\frac{3v}{1+v} + \frac{5v}{4+5v}\right)\right] v \\
&= \left[\frac{d}{dv}\left(3 - \frac{3}{1+v} + 1 - \frac{4}{4+5v}\right)\right] v \text{ (same but more conveniently written)} \\
&= \left(\frac{3}{(1+v)^2} + \frac{20}{(4+5v)^2}\right) v \\
&\approx 0.7486, \text{ since } v = e^{\beta} = 0.3145
\end{aligned}
$$

Thus,

$$
Var(\widehat{\beta}) \approx 1/0.7486 \Longrightarrow StErr(\widehat{\beta}) = 1.1558. \tag{11}
$$

Hypothesis testing: We want to test $H_0 : \beta = \beta_0 \ vs \ H_1 : \beta \neq \beta_0$, for e.g. at a 5% threshold. In particular, we often want to test $H_0 : \beta = 0 \ vs \ H_1 : \beta \neq 0$, the hypothesis of no difference between the groups (since $e^\beta = e^0 = 1$). We describe three approaches:

(a) z **test**. Compute $\widehat{\beta}$ by solving $U(\beta) = 0$ and its standard error by evaluating $SE(\widehat{\beta}) = 1/\sqrt{Inf(\widehat{\beta})}$. Refer $z = \frac{\widehat{\beta}}{SE(\widehat{\beta})}$ to $\mathcal{N}(0, 1)$ or z^2 to χ^2_1.

(b) **LRT.** Compute $-2 \log \Lambda = -2(l_0 - l_1)$ where l_i is the maximised values of l under H_i, $i = 0, 1$. Refer to χ^2_1.

(c) **Score test**. This has the advantage that is not necessary to compute $\widehat{\beta}$.The score test is

$$S = \frac{U(\beta_0)^2}{I(\beta_0)} \tag{12}$$

and has a χ^2 distribution with 1 degree of freedom if β is a scalar.

NB: All three tests are asymptotically equivalent.

The z test

We solve $U(\beta) = 0$ (graphical, Newton-Raphson, trial and error, or exactly (as in this example)). We know that $\widehat{\beta} = -1.1567$ with standard error 1.1558. Hence

$$z = \frac{-1.1567}{1.1558} = -1 \tag{13}$$

and refer to the $\mathcal{N}(0,1)$ tables, or z^2 to the χ^2_1 tables.

The LRT test

We know

$$l(\beta) = \beta - 3\log\left(1 + e^{\beta}\right) - \log\left(4 + 5e^{\beta}\right)$$

so we easily compute

$$
\begin{aligned}
l_0 &= l(0) = -3\log 2 - \log 9 = -4.276666 \qquad (14)\\
l_1 &= l(\widehat{\beta}) = -3.694984\\
-2\log\Lambda &= -2\,(-4.276666 + 3.694984) = 1.1634
\end{aligned}
$$

The score test

We have done most of the work for the score test of $H_0 : \beta = 0 \ v \ H_1 : \beta \neq 0$ already. We have with $\upsilon = \exp(\beta) = 1$ at $\beta = 0$

$$\begin{aligned} U(\beta) &= 1 - \frac{3\upsilon}{1+\upsilon} - \frac{5\upsilon}{4+5\upsilon} \qquad (15) \\ &\Longrightarrow U(0) = 1 - \frac{3}{2} - \frac{5}{9} = -\frac{19}{18} \\ Inf(\beta) &= -\frac{\partial^2 l}{\partial \beta^2} = \left(\frac{3}{(1+\upsilon)^2} + \frac{20}{(4+5\upsilon)^2} \right) \upsilon \\ &\Longrightarrow Inf(0) = \frac{3}{4} + \frac{20}{81} = \frac{323}{324} \end{aligned}$$

The score test of $H_0 : \beta = 0$ against $H_1 : \beta \neq 0$ is

$$S = \frac{U(0)^2}{Inf(0)} = \left(\frac{19}{18}\right)^2 \Big/ \left(\frac{323}{324}\right) = 1.12 \qquad (16)$$

Comment: Of course, the z test, the score test and the **LRT** are all consistent; we have $z^2 = 1.00$, $S = 1.12$ and $-2 \log \Lambda = 1.16$. None of these tests provides any sort of evidence that the mortality of the two groups differ - but then the sample size is very small.