Overview of Time-to-Event Analysis Theory and Concepts

Miguel Angel Luque-Fernandez, PhD

1. Miguel Servet I (ISCIII). Biomedical Research Institute of Granda
University of Granada, Andalusian School of Public Health
2. Honorary Professor of Epidemiology and Biostatistics
London School of Hygiene and Tropical Medicine
https://maluque.netlify.com
https://github.com/migariane/SVA-ULB

March 7, 2018

Table of Contents

Introduction

Time-to-Event

- The analysis of **(population-based cancer survival)** requires a basic understanding of the statistical methods used for **censored** survival data.

Introduction

Time-to-Event

- The analysis of **(population-based cancer survival)** requires a basic understanding of the statistical methods used for **censored** survival data.
- Survival analysis is concerned with analyzing the time to the occurrence of an event **(Time-to-Event)**.

Introduction

Time-to-Event

- The analysis of **(population-based cancer survival)** requires a basic understanding of the statistical methods used for **censored** survival data.
- Survival analysis is concerned with analyzing the time to the occurrence of an event **(Time-to-Event)**.
- The **setting of time** in your analysis is critical and it should be carefully thought (advise: study carefully the **stset** Stata command).

Introduction

Time-to-Event

- The analysis of **(population-based cancer survival)** requires a basic understanding of the statistical methods used for **censored** survival data.
- Survival analysis is concerned with analyzing the time to the occurrence of an event **(Time-to-Event)**.
- The **setting of time** in your analysis is critical and it should be carefully thought (advise: study carefully the **stset** Stata command).
- When an event occurs it is usually named a **failure**.

Time-to-Event

- The analysis of **(population-based cancer survival)** requires a basic understanding of the statistical methods used for **censored** survival data.
- Survival analysis is concerned with analyzing the time to the occurrence of an event **(Time-to-Event)**.
- The **setting of time** in your analysis is critical and it should be carefully thought (advise: study carefully the **stset** Stata command).
- When an event occurs it is usually named a **failure**.

CENSORING

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be:

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be: **Censored**

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be: **Censored**

Right censoring might occur because:

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be: **Censored**

Right censoring might occur because:

- The subject **leaves** the study before the end of the follow-up for **unknown** reasons (lost to follow-up)

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be: **Censored**

Right censoring might occur because:

- The subject **leaves** the study before the end of the follow-up for **unknown** reasons (lost to follow-up)
- The subject **withdraws** prematurely (e.g., drug intolerance in a clinical trial)

Censoring

censoring

In practice, you often do not know when failures occurred, at least not for every observation in the data set.

When it happen the observation is said to be: **Censored**

Right censoring might occur because:

- The subject **leaves** the study before the end of the follow-up for **unknown** reasons (lost to follow-up)
- The subject **withdraws** prematurely (e.g., drug intolerance in a clinical trial)

Informative Censoring

Informative censoring

- For the majority of analytical tools used for time-to-event analysis, we assume that if censoring occurs, it **OCCURS RANDOMLY** (i.e. **unrelated** to the reason for failure)

Informative Censoring

Informative censoring

- For the majority of analytical tools used for time-to-event analysis, we assume that if censoring occurs, it **OCCURS RANDOMLY** (i.e. **unrelated** to the reason for failure)

- Hence, **informative censoring** occurs when the underlying cause of withdrawal it is associated with the outcome or failure

Informative Censoring

Informative censoring

- For the majority of analytical tools used for time-to-event analysis, we assume that if censoring occurs, it **OCCURS RANDOMLY** (i.e. **unrelated** to the reason for failure)

- Hence, **informative censoring** occurs when the underlying cause of withdrawal it is associated with the outcome or failure

- Any example**?**

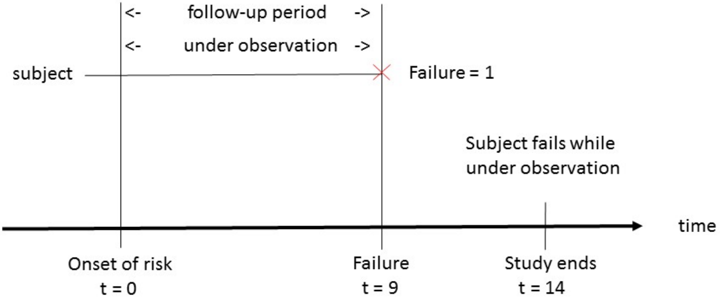
Informative Censoring

Informative censoring

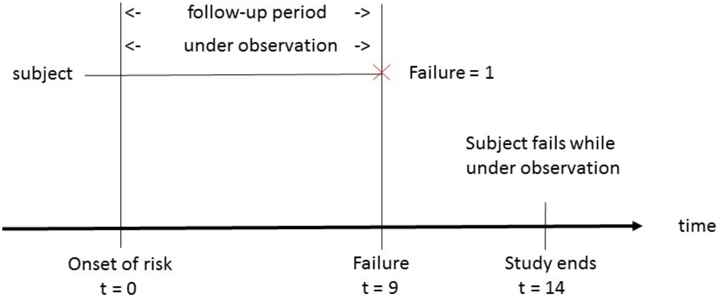
- For the majority of analytical tools used for time-to-event analysis, we assume that if censoring occurs, it **OCCURS RANDOMLY** (i.e. **unrelated** to the reason for failure)
- Hence, **informative censoring** occurs when the underlying cause of withdrawal it is associated with the outcome or failure
- Any example**?**

Let's practice! It is your turn!

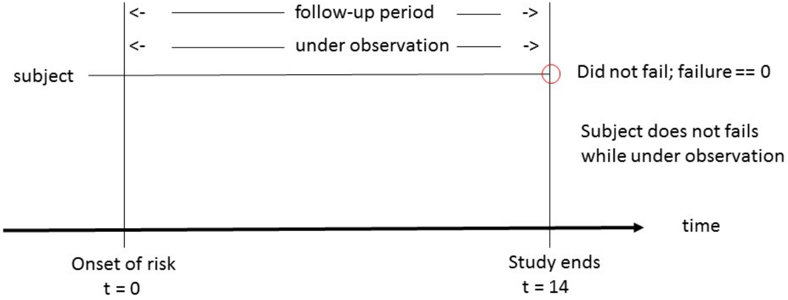
It is censoring?



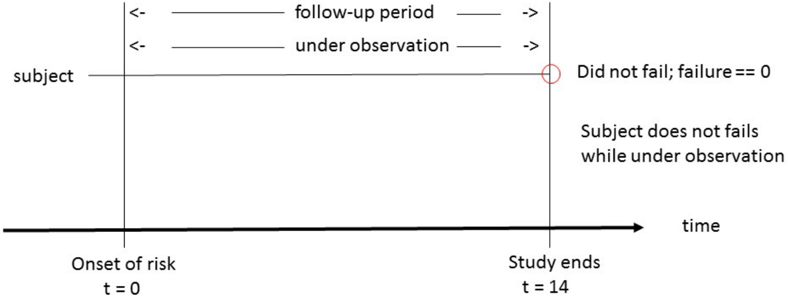
No, it is just a failure (the event of interest)



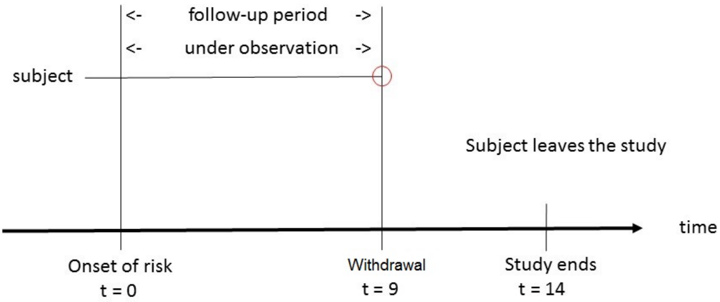
It is censoring?



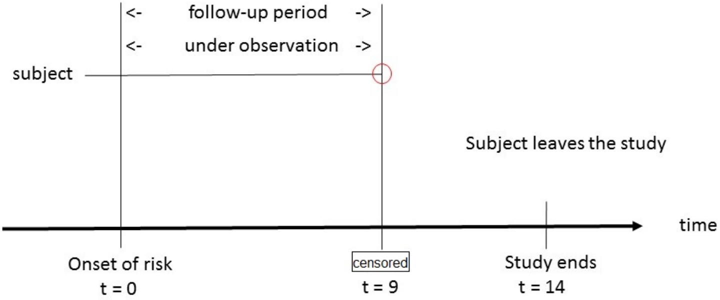
Yes, right censoring at the end of the study



It is censoring?



Yes, right censoring before the end of the study (LTFU)



TRUNCATION

Truncation

Truncation concept

- **Truncation** is often confused with censoring because it also gives rise to incomplete observations over time

Truncation

Truncation concept

- **Truncation** is often confused with censoring because it also gives rise to incomplete observations over time
- **Truncation**, in most statistical applications, refers to **complete ignorance** about the variable of interest and the covariates over a portion of the distribution

Truncation

Truncation concept

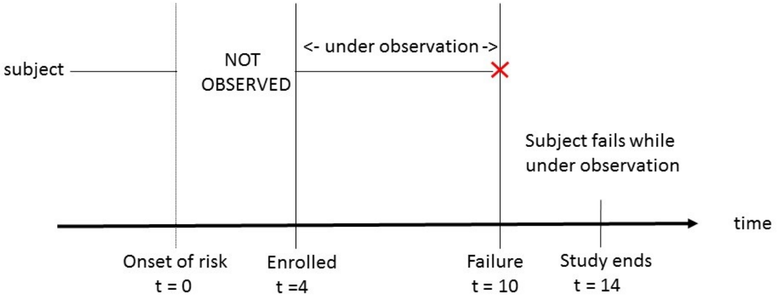
- **Truncation** is often confused with censoring because it also gives rise to incomplete observations over time
- **Truncation**, in most statistical applications, refers to **complete ignorance** about the variable of interest and the covariates over a portion of the distribution
- **Left Truncation** is used in PERIOD analysis.

Truncation

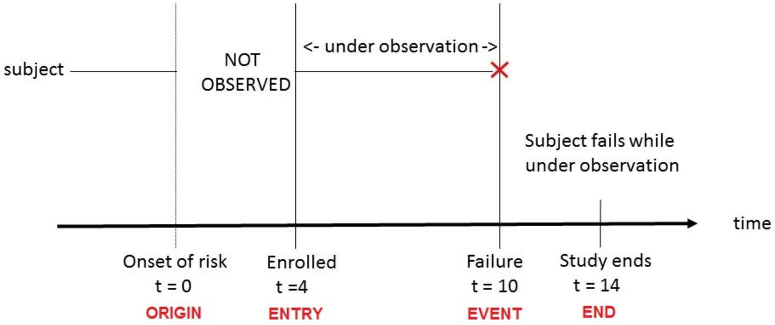
Truncation concept

- **Truncation** is often confused with censoring because it also gives rise to incomplete observations over time
- **Truncation**, in most statistical applications, refers to **complete ignorance** about the variable of interest and the covariates over a portion of the distribution
- **Left Truncation** is used in PERIOD analysis.

Left Truncation

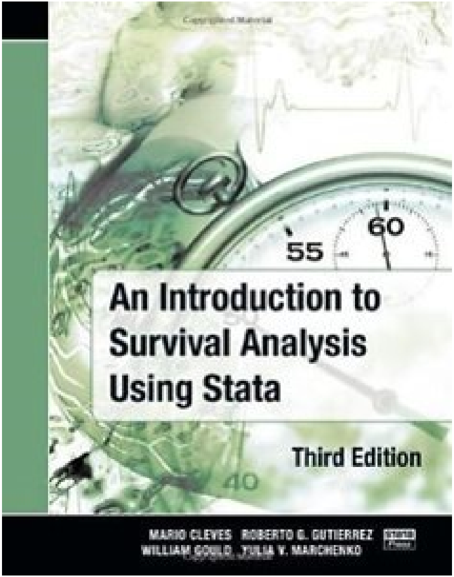


Left Truncation



Setting TIME in Statistical Software

Setting of time: chapters 5,6,7



Stata stset command

stset finmdy, fail(dead==1) origin(time diagmdy) enter(time diagmdy)

```
     failure event:  dead == 1
obs. time interval:  (origin, finmdy]
 enter on or after:  time diagmdy
 exit on or before:  failure
    t for analysis:  (time-origin)
            origin:  time diagmdy
------------------------------------------------------------------------------
     355801  total observations
          0  exclusions
------------------------------------------------------------------------------
     355801  observations remaining, representing
     212121  failures in single-record/single-failure data
  920269050  total analysis time at risk and under observation
                                                at risk from t =         0
                                     earliest observed entry t =         0
                                          last observed exit t =     12052
```

Listing the output of the Stata stset command

list diagmdy finmdy _t0 _t _d _st in 1/10

```
     +-----------------------------------------------------+
     |   diagmdy      finmdy   _t0          _t   _d    _st |
     |-----------------------------------------------------|
  1. | 01jan2000   31dec2003     0   3.9972621    1      1 |
  2. | 29jul1998   31dec2003     0   5.4236824    0      1 |
  3. | 30jan1998   31dec2003     0   5.9164956    0      1 |
  4. | 09jul1998   31dec2003     0   5.4784394    0      1 |
  5. | 22dec1998   31dec2003     0   5.0239562    0      1 |
     |-----------------------------------------------------|
  6. | 23jan1998   31dec2003     0   5.9356605    0      1 |
  7. | 16jul1998   31dec2003     0   5.4592745    0      1 |
  8. | 07jul1999   31dec2003     0   4.4845996    0      1 |
  9. | 18aug1998   31dec2003     0   5.3689254    0      1 |
 10. | 01sep1999   01oct2003     0   4.0821355    1      1 |
     +-----------------------------------------------------+
```

Origin and Entry

Important

It is critical to distinguish between the **time** of the **STUDY ENTRY** and the **ORIGIN** of **time.**

Time

- **origin:** It defines **time 0**, that is, when we start **recording time** (date of birth, date of cancer diagnosis, etc).

Setting time in Stata: concepts

Time

- **origin:** It defines **time 0**, that is, when we start **recording time** (date of birth, date of cancer diagnosis, etc).
- **Note:** Time "0" in Stata is **1.1.1960**.

Setting time in Stata: concepts

Time

- **origin:** It defines **time 0**, that is, when we start **recording time** (date of birth, date of cancer diagnosis, etc).

- **Note:** Time ”0” in Stata is **1.1.1960**.

- **enter:** When the subject enter the study (date) and **starts being at risk** (in many cases, this is the same as the time origin: data of cancer diagnosis)

Setting time in Stata: concepts

Time

- **origin:** It defines **time 0**, that is, when we start **recording time** (date of birth, date of cancer diagnosis, etc).
- **Note:** Time ”0” in Stata is **1.1.1960**.
- **enter:** When the subject enter the study (date) and **starts being at risk** (in many cases, this is the same as the time origin: data of cancer diagnosis)
- **exit:** The time (or date) when the subject leaves the study and **stops being at risk** either by experiencing the event or being censored

Setting time in Stata: concepts

Time

- **origin:** It defines **time 0**, that is, when we start **recording time** (date of birth, date of cancer diagnosis, etc).
- **Note:** Time ”0” in Stata is **1.1.1960**.
- **enter:** When the subject enter the study (date) and **starts being at risk** (in many cases, this is the same as the time origin: data of cancer diagnosis)
- **exit:** The time (or date) when the subject leaves the study and **stops being at risk** either by experiencing the event or being censored

Setting time in Stata: concepts

Time

- _t0 = Records the start of the study,

Setting time in Stata: concepts

Time

- **_t0** = Records the start of the study,
- **_t** = Records the end of the study

Setting time in Stata: concepts

Time

- **_t0** = Records the start of the study,
- **_t** = Records the end of the study
- **_d**: Records the outcome at the end of the span of the study (0, 1).

Setting time in Stata: concepts

Time

- **_t0** = Records the start of the study,
- **_t** = Records the end of the study
- **_d**: Records the outcome at the end of the span of the study (0, 1).
- **_st**: Records whether the observation is going to be used in the current analysis

Setting time in Stata: concepts

Time

- **_t0** = Records the start of the study,
- **_t** = Records the end of the study
- **_d**: Records the outcome at the end of the span of the study (0, 1).
- **_st**: Records whether the observation is going to be used in the current analysis
- **scale(1)** as default but scale(30.25), scale(365.25), (...).

Setting time in Stata: concepts

Time

- **_t0** = Records the start of the study,
- **_t** = Records the end of the study
- **_d**: Records the outcome at the end of the span of the study (0, 1).
- **_st**: Records whether the observation is going to be used in the current analysis
- **scale(1)** as default but scale(30.25), scale(365.25), (...).
  - $\frac{\textbf{time} - \textbf{origin(0)}}{\textbf{scale(1)}}$

Setting time in Stata: concepts

Time

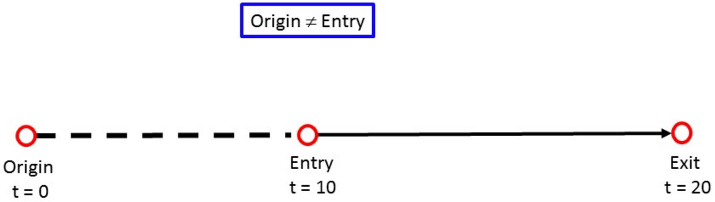
- **_t0** = Records the start of the study,
- **_t** = Records the end of the study
- **_d**: Records the outcome at the end of the span of the study (0, 1).
- **_st**: Records whether the observation is going to be used in the current analysis
- **scale(1)** as default but scale(30.25), scale(365.25), (...).
  - $\frac{\textbf{time} - \textbf{origin(0)}}{\textbf{scale(1)}}$

Study entry and Time Origin

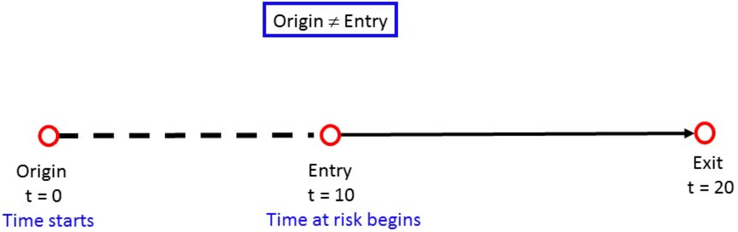
Origin $\not\approx$ Entry

Let's practice! It is your turn!

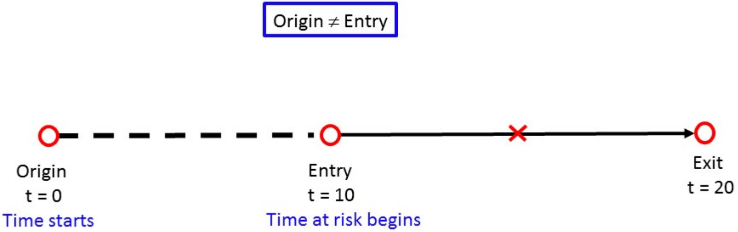
Origin and Study Entry



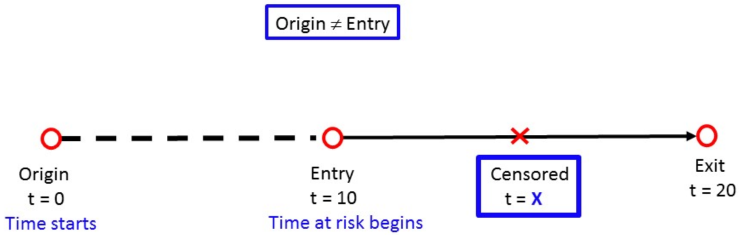
Origin and Study Entry



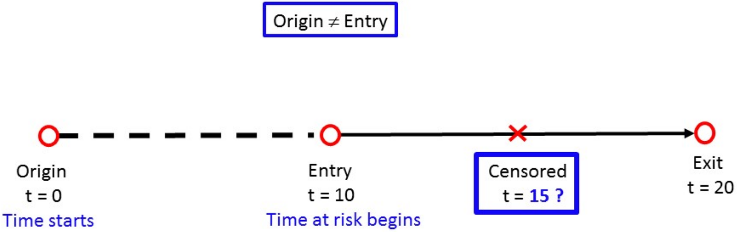
Origin and Study Entry



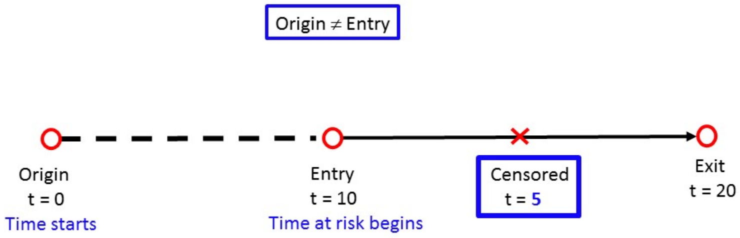
Origin and Study Entry



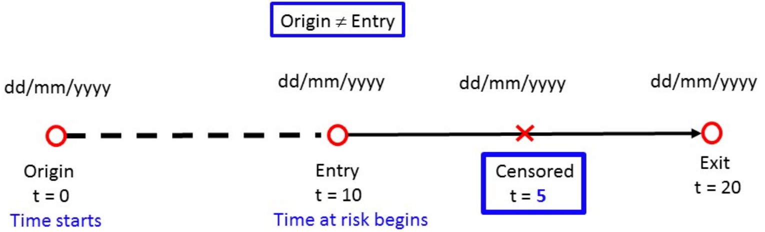
Origin and Study Entry



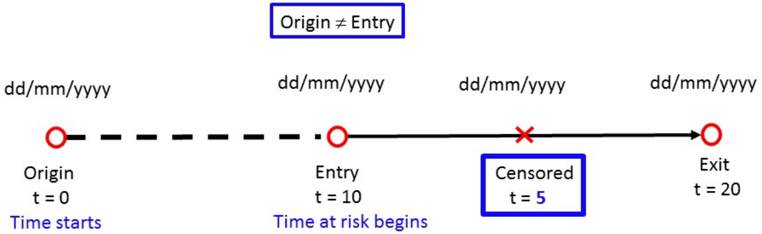
Origin and Study Entry



Origin and Study Entry

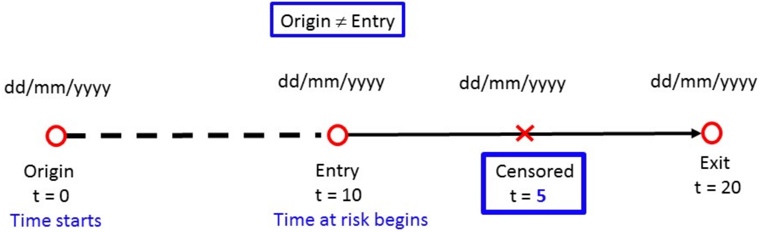


$$\text{scale} = \frac{\text{exit(dd/mm/yy)} - \text{origin(dd/mm/yy)}}{\text{scale}(365.25)}$$



$$\textbf{scale} = \frac{\text{exit(dd/mm/yy)} - \textbf{origin}\text{(dd/mm/yy)}}{\text{scale}(365.25)} \neq \textbf{person-time at risk} = \frac{\text{exit(dd/mm/yy)} - \textbf{entry}\text{(dd/mm/yy)}}{\text{scale}(365.25)}$$

Time zero in Stata = **01.JAN.1960**



$$\text{scale} = \frac{\text{exit(dd/mm/yy)} - \textbf{origin}\text{(dd/mm/yy)}}{\text{scale(365.25)}} \neq \text{person-time at risk} = \frac{\text{exit(dd/mm/yy)} - \textbf{entry}\text{(dd/mm/yy)}}{\text{scale(365.25)}}$$

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**
- **stvary** for multiple records, for which id() has been stset

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**
- **stvary** for multiple records, for which id() has been stset
- **sts graph, gwood** (Kaplan-Meier curve 95%CI)

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**
- **stvary** for multiple records, for which id() has been stset
- **sts graph, gwood** (Kaplan-Meier curve 95%CI)
- **sts list** (Life Table)

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**
- **stvary** for multiple records, for which id() has been stset
- **sts graph, gwood** (Kaplan-Meier curve 95%CI)
- **sts list** (Life Table)
- **stci** (median survival time)

Useful Stata commands after stset

Describing your Cohort

- **stdescribe**
- **stvary** for multiple records, for which id() has been stset
- **sts graph, gwood** (Kaplan-Meier curve 95%CI)
- **sts list** (Life Table)
- **stci** (median survival time)

Useful Stata commands after stset

Describing your Cohort

- **stci p(#)** (percentiles of the survival time)

Useful Stata commands after stset

Describing your Cohort

- **stci p(#)** (percentiles of the survival time)
- **stsum, by(var)** (incidence rate, person-time at risk, # subjects)

Useful Stata commands after stset

Describing your Cohort

- **stci p(#)** (percentiles of the survival time)
- **stsum, by(var)** (incidence rate, person-time at risk, # subjects)
- **stci, rmean by(var)** (Restricted Mean Survival Time (RMST), SE, 95%CI)

Useful Stata commands after stset

Describing your Cohort

- **stci p(#)** (percentiles of the survival time)
- **stsum, by(var)** (incidence rate, person-time at risk, # subjects)
- **stci, rmean by(var)** (Restricted Mean Survival Time (RMST), SE, 95%CI)
- **sts test** (Log-rank test)

Useful Stata commands after stset

Describing your Cohort

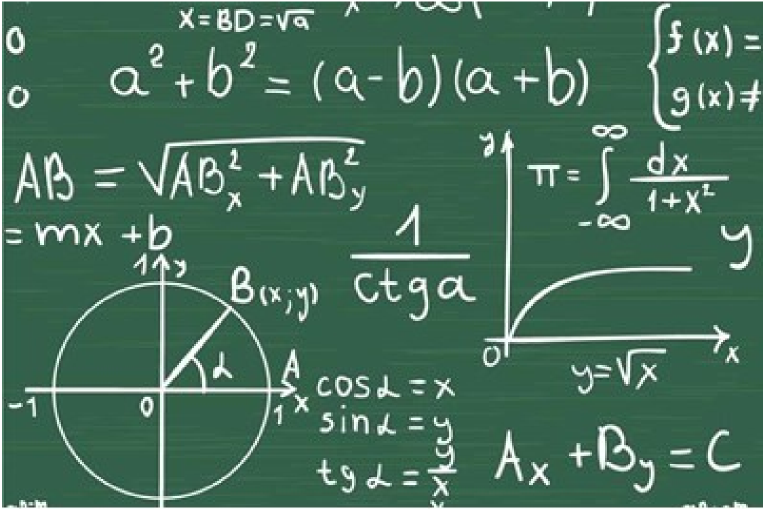
- **stci p(#)** (percentiles of the survival time)
- **stsum, by(var)** (incidence rate, person-time at risk, # subjects)
- **stci, rmean by(var)** (Restricted Mean Survival Time (RMST), SE, 95%CI)
- **sts test** (Log-rank test)

The Maths behind Survival

The anatomy of time-to-event analysis

Anatomische les door Dokter van der Meer, 1617

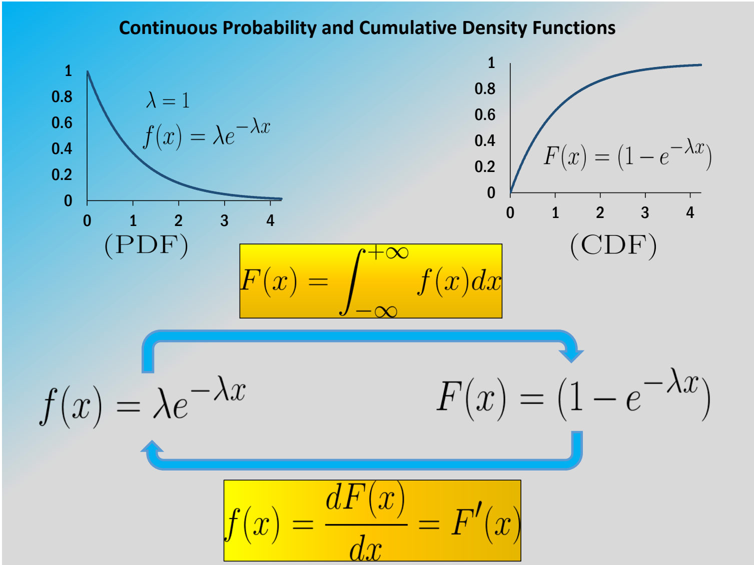
Epidemiologists



The Maths behind Survival

The Probability Density Function PDF and Cumulative Density Function CDF of a random variable

PDF and CDF



Continuous Probability and Cumulative Density Functions
1
0.8
0.6
0.4
0.2
0
0 1 2 3 4
$\lambda = 1$
$f(x) = \lambda e^{-\lambda x}$
(PDF)
$F(x) = (1 - e^{-\lambda x})$
(CDF)
$F(x) = \int_{-\infty}^{+\infty} f(x)dx$
$f(x) = \lambda e^{-\lambda x}$
$F(x) = (1 - e^{-\lambda x})$
$f(x) = \frac{dF(x)}{dx} = F'(x)$

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Now we can readily derive:

- the instantaneous hazard h(t),

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Now we can readily derive:

- the instantaneous hazard h(t),
- the cumulative hazard H(t),

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Now we can readily derive:

- the instantaneous hazard h(t),
- the cumulative hazard H(t),
- the survival S(t),

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Now we can readily derive:

- the instantaneous hazard h(t),
- the cumulative hazard H(t),
- the survival S(t),
- and their relationship easily (see next)

Derivative and antiderivative

Fundamental theorem to memorize

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

Now we can readily derive:

- the instantaneous hazard h(t),
- the cumulative hazard H(t),
- the survival S(t),
- and their relationship easily (see next)

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\text{\#events(failures)}}{\text{\#survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}} \text{du} =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} = \int_0^t \frac{1}{\text{S(u)}}\text{(1 - S(u))'} =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} = \int_0^t \frac{1}{\text{S(u)}}\text{(1 - S(u))'} = \text{-log(S(t))}$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} = \int_0^t \frac{1}{\text{S(u)}}\text{(1 - S(u))'} = \text{-log(S(t))}$$

h(t): Instantaneous hazard or λ

$$\lambda(t) =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} = \int_0^t \frac{1}{\text{S(u)}}\text{(1 - S(u))'} = \text{-log(S(t))}$$

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\text{f(t)}}{\text{S(t)}} =$$

PDF and CDF: time-to-event (Failure)

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\#\text{events(failures)}}{\#\text{survivors}} = \text{P(event at t} \mid \text{survival up to t)} = \frac{\text{f(t)}}{\text{S(t)}}$$

H(t): Cumulative hazard or Λ

$$\text{H(t)} = \int_0^t \frac{\text{f(u)}}{\text{S(u)}}\text{du} = \int_0^t \frac{1}{\text{S(u)}}\text{(F')} = \int_0^t \frac{1}{\text{S(u)}}\text{(1 - S(u))'} = \text{-log(S(t))}$$

h(t): Instantaneous hazard or λ

$$\lambda(t) = \frac{\text{f(t)}}{\text{S(t)}} = \frac{\text{-dlog(S(t))}}{\text{dt}}$$

The soul of Survival Analysis

From the h(t) we can derive:

$$H(t) =$$

The soul of Survival Analysis

From the h(t) we can derive:

$$H(t) = -\log(S(t))$$

From the h(t) we can derive:

$$H(t) = -\log(S(t))$$

$$\exp(H(t)) = -S(t)$$

The soul of Survival Analysis

From the h(t) we can derive:

$$H(t) = -\log(S(t))$$

$$\exp(H(t)) = -S(t)$$

$$S(t) = \exp(-H(t))$$

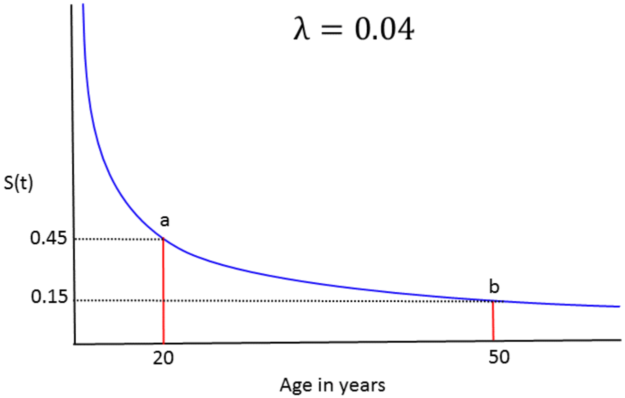
Formulae to derive h(t), H(t) and S(t): Pintilie M.(2007)

Table 2.1 Basic mathematical formulae.

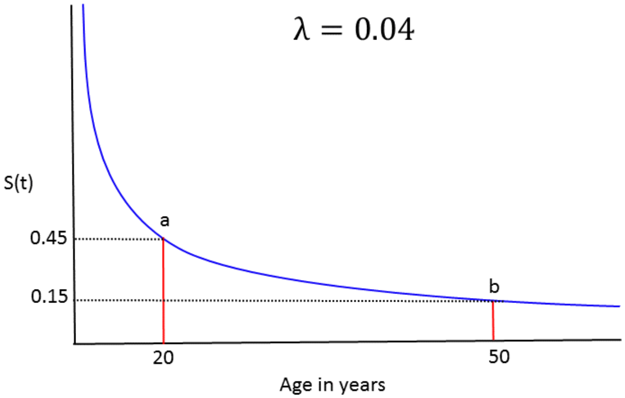
	Continuous distributions		Discrete distributions	
Function	Definition	Relationship to other functions	Definition	Relationship to other functions
Cumulative distribution	$F(t) = P(T \le t)$	$F(t) = \int_0^t f(x)dx$	$F(t) = P(T \le t)$	$F(t) = \sum_{t_j \le t} p(t_j)$
Probability density	$f(t) = \frac{d}{dt}F(t)$	$f(t) = -\frac{d}{dt}S(t)$	$p(t_j) = P(T = t_j)$	$p(t_j) = S(t_{j-1}) - S(t_j)$
Survivor	$S(t) = P(T > t)$	$S(t) = 1 - F(t)$ $= \int_t^\infty f(x)dx$ $= \exp\{-H(t)\}$	$S(t) = P(T > t)^*$	$S(t) = 1 - F(t)$ $= \sum_{t_j > t} p(t_j)$ $= \exp\{-H_2(t)\}$
Hazard	$h(t) = \lim_{\delta t \to 0} \left\{ \frac{P(t \le T < t + \delta t \mid T > t)}{\delta t} \right\}$	$h(t) = f(t)/S(t)$ $= -\frac{d}{dt} \log S(t)$	$h(t_j) = P(T = t_j \mid T \ge t_j)$ $H_1(t) = \sum_{t_j \le t} h(t_j)$	$h(t_j) = p(t_j)/S(t_{j-1})$
Cumulative hazard	$H(t) = \int_0^t h(x)dx$	$H(t) = -\log S(t)$	$H_2(t) = -\sum_{t_j \le t} \log\{1 - h(t_j)\}$	$H_2(t) = -\log S(t)$

* Probability function

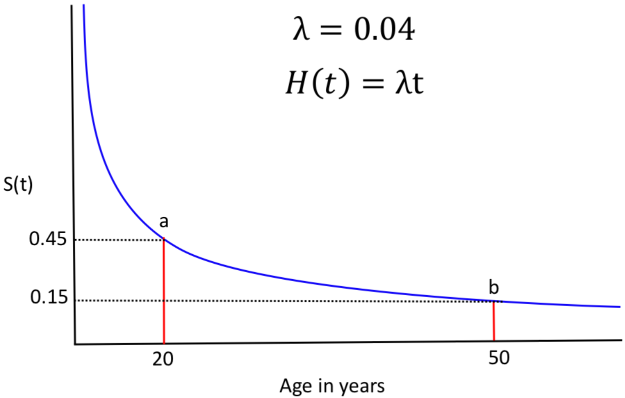
Deriving h(t), H(t), and S(t) from the Exponential distribution



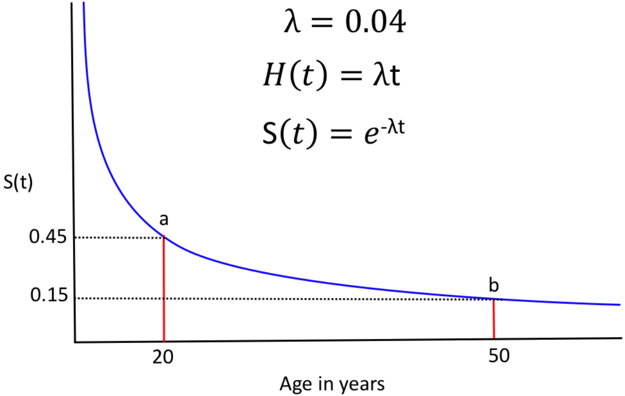
Calculus and Geometry



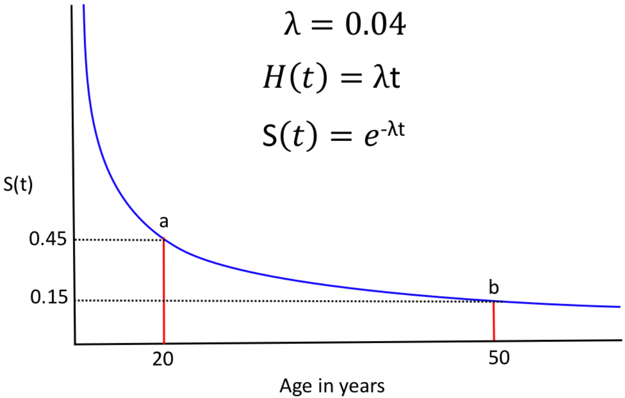
Calculus and Geometry



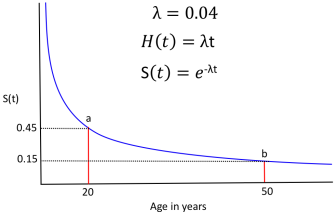
Calculus and Geometry



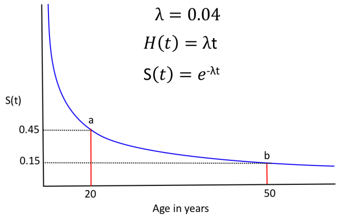
Calculus and Geometry



Calculus and Geometry

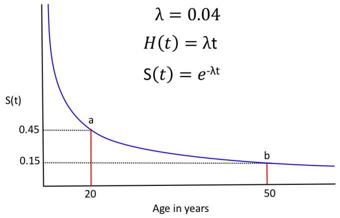


Calculus and Geometry



Link between hazard and velocity

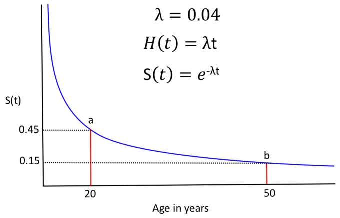
Calculus and Geometry



Link between hazard and velocity

- Velocity in Km/h $\approx$ Instantaneous hazard (λ) in person-years

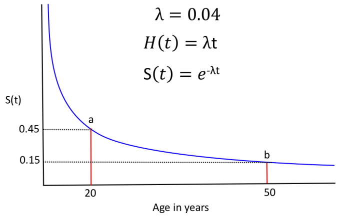
Calculus and Geometry



Link between hazard and velocity

- Velocity in Km/h $\approx$ Instantaneous hazard (λ) in person-years
- You can compute the **average speed** from point **a** to **b** in **Km**

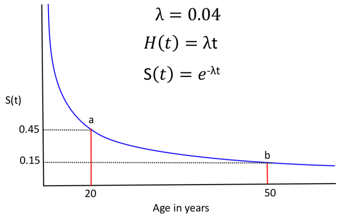
Calculus and Geometry



Link between hazard and velocity

- Velocity in Km/h $\approx$ Instantaneous hazard (λ) in person-years
- You can compute the **average speed** from point **a** to **b** in **Km**
- In the same way you can compute the **time** from time **a** to **b** in **years**

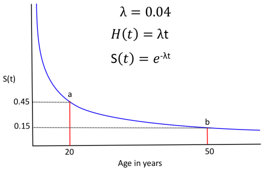
Calculus and Geometry



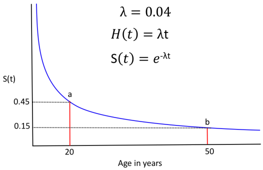
Link between hazard and velocity

- Velocity in Km/h $\approx$ Instantaneous hazard (λ) in person-years
- You can compute the **average speed** from point **a** to **b** in **Km**
- In the same way you can compute the **time** from time **a** to **b** in **years**

Calculus and Geometry

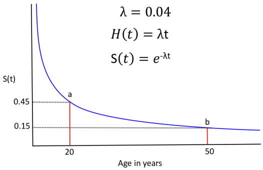


Calculus and Geometry



$$H(t) = \lambda t \, ; t = \frac{H(t)}{\lambda} \, ; \lambda = \frac{H(t)}{t}$$

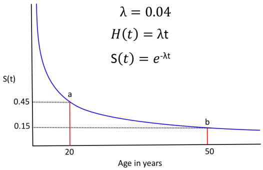
Calculus and Geometry



$$H(t) = \lambda t \; ; \; t = \frac{H(t)}{\lambda} \; ; \; \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b (\text{1- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

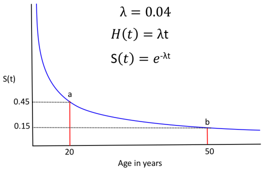


$$H(t) = \lambda t \,;\, t = \frac{H(t)}{\lambda} \,;\, \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b \text{(1- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

$\lambda =$



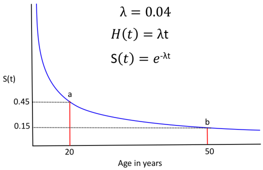
$$H(t) = \lambda t \,;\, t = \frac{H(t)}{\lambda} \,;\, \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b \text{(1- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

$$\lambda = \frac{\text{H(t)}}{\text{t}} =$$

Calculus and Geometry



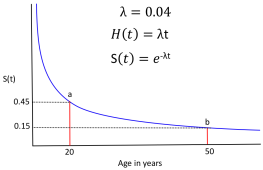
$$H(t) = \lambda t \; ; t = \frac{H(t)}{\lambda} \; ; \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b (1\text{- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

$$\lambda = \frac{\text{H(t)}}{\text{t}} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{7.85} =$$

Calculus and Geometry



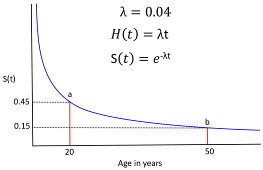
$$H(t) = \lambda t \; ; t = \frac{H(t)}{\lambda} \; ; \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b \text{(1- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

$$\lambda = \frac{\text{H(t)}}{\text{t}} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{7.85} = 0.04 = 4 \text{ events per 100 person-year}$$

Calculus and Geometry



$$H(t) = \lambda t \; ; t = \frac{H(t)}{\lambda} \; ; \lambda = \frac{H(t)}{t}$$

$$\text{Time} = \frac{\int_a^b (1\text{- P(T} > \text{b) - (1-P(T} > \text{a)}}{\lambda} = \frac{(1 - e^{-\lambda b})(1 - e^{-\lambda a})}{\lambda} =$$

$$\frac{(e^{-\lambda a} - e^{-\lambda b})}{\lambda} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{0.04} = 7.85 \text{ years}$$

$$\lambda = \frac{\text{H(t)}}{\text{t}} = \frac{(e^{-0.04*20} - e^{-0.04*50})}{7.85} = 0.04 = 4 \text{ events per 100 person-year}$$

The Statistical Methods

The Statistical Methods used in time-to-event analysis

Methods

Statistical Methods

Non-parametric

Methods

Statistical Methods

Non-parametric

- Life Tables

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull
- Log-normal

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull
- Log-normal
- Generalized Gamma

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull
- Log-normal
- Generalized Gamma

Semiparametric Cox Regression Model

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull
- Log-normal
- Generalized Gamma

Semiparametric Cox Regression Model

Flexible parametric methods

Methods

Statistical Methods

Non-parametric

- Life Tables
- Kaplan-Meier
- Nelson-Aalen

Parametric

- Exponential
- Weibull
- Log-normal
- Generalized Gamma

Semiparametric Cox Regression Model

Flexible parametric methods

The Statistical Methods

The Non-parametric Methods used in time-to-event analysis

Life-Table

Statistical Methods

- **John Graunt** **Edmond Halley** Demography and Actuarial Science.

Life-Table

Statistical Methods

- **John Graunt Edmond Halley** Demography and Actuarial Science.
- LT enables us to build up the **survival experience** of a cohort over a period of time

Life-Table

Statistical Methods

- **John Graunt Edmond Halley** Demography and Actuarial Science.
- LT enables us to build up the **survival experience** of a cohort over a period of time
- LT breaks down the total follow-up into shorter consecutive intervals **(discrete time)**

Life-Table

Statistical Methods

- **John Graunt Edmond Halley** Demography and Actuarial Science.
- LT enables us to build up the **survival experience** of a cohort over a period of time
- LT breaks down the total follow-up into shorter consecutive intervals **(discrete time)**
- Population-based LTs are used to estimate **RELATIVE SURVIVAL**

Life-Table

Statistical Methods

- **John Graunt Edmond Halley** Demography and Actuarial Science.
- LT enables us to build up the **survival experience** of a cohort over a period of time
- LT breaks down the total follow-up into shorter consecutive intervals **(discrete time)**
- Population-based LTs are used to estimate **RELATIVE SURVIVAL**
- The population **MORTALITY RATE or HAZARD** is crucial and refers to the **background mortality experience**

Life-Table

Statistical Methods

- **John Graunt Edmond Halley** Demography and Actuarial Science.
- LT enables us to build up the **survival experience** of a cohort over a period of time
- LT breaks down the total follow-up into shorter consecutive intervals **(discrete time)**
- Population-based LTs are used to estimate **RELATIVE SURVIVAL**
- The population **MORTALITY RATE or HAZARD** is crucial and refers to the **background mortality experience**

Table 1. Life Table Actuarial Method

						Probability	
	Total a_i	Deaths d_i	Lost c_i	N.A.R $n_i = (a_i - (c_i/2))$	Rate $r_i = d_i/n_i$	s(t)$= 1 - r_i$	S(t)=$S(i-1) \times S_i$
00-01	100	35	0	100	0.350	0.650	0.650
01-02	65	6	10	60	0.100	0.900	0.585
02-03	49	5	18	40	0.125	0.875	0.512
03-04	26	1	7	22.5	0.0444	0.955	0.489

Table 1. Life Table Actuarial Method

	Total a_i	Deaths d_i	Lost c_i	N.A.R $n_i = (a_i - (c_i/2))$	Rate $r_i = d_i/n_i$	Probability s(t)$= 1 - r_i$	S(t)=$S(i-1) \times S_i$
00-01	100	35	0	100	0.350	0.650	0.650
01-02	65	6	10	60	0.100	0.900	0.585
02-03	49	5	18	40	0.125	0.875	0.512
03-04	26	1	7	22.5	0.0444	0.955	0.489

Life Table in Stata

```
In Stata use sts list
            Beg.              Net             Survivor      Std.
  Time      Total    Fail    Lost             Function     Error      [95% Conf. Int.]
-------------------------------------------------------------------------------
   1-6        47       8       0               0.8298     0.0548     0.6883    0.9110
  6-11        39       7       4               0.6809     0.0680     0.5275    0.7937
 11-16        28       6       2               0.5350     0.0751     0.3786    0.6684
 (...)
 31-39         5       1       4               0.1887     0.0737     0.0714    0.3486
-------------------------------------------------------------------------------
```

Table 1. Life Table Actuarial Method

	Total a_i	Deaths d_i	Lost c_i	**N.A.R** $n_i = (a_i - (c_i/2))$	Rate $r_i = d_i/n_i$	Probability s(t)$= 1 - r_i$	S(t)=$S(i-1) \times S_i$
00-01	100	35	0	100	0.350	0.650	0.650
01-02	65	6	10	60	0.100	0.900	0.585
02-03	49	5	18	40	0.125	0.875	0.512
03-04	26	1	7	22.5	0.0444	0.955	0.489

Life Table in Stata

```
In Stata use sts list
           Beg.             Net              Survivor      Std.
  Time     Total    Fail    Lost             Function      Error      [95% Conf. Int.]
-------------------------------------------------------------------------------------
   1-6       47       8      0                0.8298     0.0548     0.6883     0.9110
  6-11       39       7      4                0.6809     0.0680     0.5275     0.7937
 11-16       28       6      2                0.5350     0.0751     0.3786     0.6684
 (...)
 31-39        5       1      4                0.1887     0.0737     0.0714     0.3486
-------------------------------------------------------------------------------------
```

Table 1. Life Table Actuarial Method

	Total a_i	Deaths d_i	Lost c_i	**N.A.R** $n_i = (a_i - (c_i/2))$	Rate $r_i = d_i/n_i$	Probability s(t)$= 1 - r_i$	S(t)=$S(i-1) \times S_i$
00-01	100	35	0	100	0.350	0.650	**0.650**
01-02	65	6	10	60	0.100	0.900	0.585
02-03	49	5	18	40	0.125	0.875	0.512
03-04	26	1	7	22.5	0.0444	0.955	0.489

Life Table in Stata

```
In Stata use sts list
           Beg.           Net              Survivor     Std.
  Time     Total    Fail  Lost             Function     Error      [95% Conf. Int.]
--------------------------------------------------------------------------------
   1-6       47       8      0               0.8298    0.0548      0.6883    0.9110
  6-11       39       7      4               0.6809    0.0680      0.5275    0.7937
 11-16       28       6      2               0.5350    0.0751      0.3786    0.6684
 (...)
 31-39        5       1      4               0.1887    0.0737      0.0714    0.3486
--------------------------------------------------------------------------------
```

Table 1. Life Table Actuarial Method

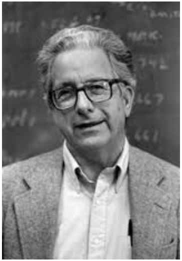
						Probability	
	Total a_i	Deaths d_i	Lost c_i	**N.A.R** $n_i = (a_i - (c_i/2))$	Rate $r_i = d_i/n_i$	s(t)$= 1 - r_i$	S(t)=$S(i-1) \times S_i$
00-01	100	35	0	100	0.350	0.650	**0.650**
01-02	65	6	10	60	0.100	**0.900**	0.585
02-03	49	5	18	40	0.125	0.875	0.512
03-04	26	1	7	22.5	0.0444	0.955	0.489

Life Table in Stata

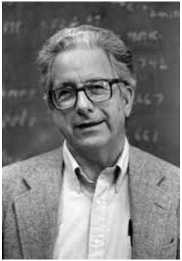
```
In Stata use sts list
          Beg.          Net            Survivor      Std.
  Time    Total   Fail  Lost           Function      Error     [95% Conf. Int.]
-------------------------------------------------------------------------------
   1-6      47      8     0              0.8298    0.0548     0.6883     0.9110
  6-11      39      7     4              0.6809    0.0680     0.5275     0.7937
 11-16      28      6     2              0.5350    0.0751     0.3786     0.6684
 (...)
 31-39       5      1     4              0.1887    0.0737     0.0714     0.3486
-------------------------------------------------------------------------------
```

Non-parametric: Edward Kaplan and Paul Meier, 1958

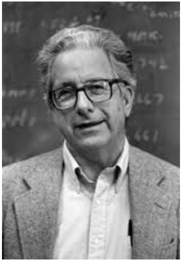
Non-parametric: Edward Kaplan and Paul Meier, 1958



Non-parametric: Edward Kaplan and Paul Meier, 1958

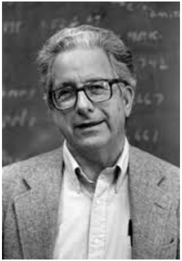


Non-parametric: Edward Kaplan and Paul Meier, 1958



No assumptions about the functional form of the baseline hazard

Non-parametric: Edward Kaplan and Paul Meier, 1958



No assumptions about the functional form of the baseline hazard

Non-parametric

Kaplan-Meier or Product Limit Estimator

$$\widehat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Non-parametric

Kaplan-Meier or Product Limit Estimator

$$\widehat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Greenwood's variance estimator

$$\widehat{Var}\left(\hat{S}(t)\right) = \widehat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

s

Non-parametric

Kaplan-Meier or Product Limit Estimator

$$\widehat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Greenwood's variance estimator

$$\widehat{Var}\left(\hat{S}(t)\right) = \widehat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

s

Confidence Intervals

$$95\% CI = \widehat{S}(t) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\widehat{S}(t))}$$

Non-parametric

Kaplan-Meier or Product Limit Estimator

$$\widehat{S}(t) = \prod_{t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Greenwood′s variance estimator

$$\widehat{Var}\left(\hat{S}(t)\right) = \widehat{S}(t)^2 \sum_{t_j \leq t} \frac{d_j}{n_j(n_j - d_j)}$$

s

Confidence Intervals

$$95\% CI = \widehat{S}(t) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{Var}(\widehat{S}(t))}$$

Kaplan-Meier by group

Table 2. Kaplan-Meier survival function by treatment group

Group	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Placebo	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Placebo	1	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Placebo	2	4	1	0	0.7500	0.6000	0.0833	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	7	3	0	1	1.0000	0.6000	0.0000	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	8	2	1	0	0.5000	0.3000	0.5000	0.6333	0.1856	0.6610	-1.1099	1.4812	0.0123	0.7192
Placebo	12	1	1	0	0.0000	0.0000								
	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Drug	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Drug	4	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	6	4	0	1	1.0000	0.8000	0.0000	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	9	3	1	0	0.6667	0.5333	0.1667	0.2167	-0.4642	0.7405	-1.9156	0.9871	0.0683	0.8631
Drug	10	2	1	0	0.5000	0.2667	0.5000	0.7167	0.2790	0.6405	-0.9764	1.5343	0.0097	0.6861
Drug	13	1	1	0	0.0000	0.0000								

Kaplan-Meier by group

Table 2. Kaplan-Meier survival function by treatment group

Group	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Placebo	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Placebo	1	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Placebo	2	4	1	0	0.7500	0.6000	0.0833	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	7	3	0	1	1.0000	0.6000	0.0000	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	8	2	1	0	0.5000	0.3000	0.5000	0.6333	0.1856	0.6610	-1.1099	1.4812	0.0123	0.7192
Placebo	12	1	1	0	0.0000	0.0000								
	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Drug	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Drug	4	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	6	4	0	1	1.0000	0.8000	0.0000	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	9	3	1	0	0.6667	0.5333	0.1667	0.2167	-0.4642	0.7405	-1.9156	0.9871	0.0683	0.8631
Drug	10	2	1	0	0.5000	0.2667	0.5000	0.7167	0.2790	0.6405	-0.9764	1.5343	0.0097	0.6861
Drug	13	1	1	0	0.0000	0.0000								

Log-Log SE based on Delta Method

$$SE = \left[ln\left[-ln(\widehat{S}(t))\right]\right]$$

$$= \sqrt{\frac{1}{\left[ln(\widehat{S}(t))\right]^2} \sum \frac{d_i}{n_i(n_i - d_i)}}$$

$$= exp(-exp(ln(-ln(\widehat{S}(t))))) \pm 1.96 \times SE(ln(-ln(\widehat{S}(t))))$$

Kaplan-Meier by group

Table 2. Kaplan-Meier survival function by treatment group

Group	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Placebo	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Placebo	1	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Placebo	2	4	1	0	**0.7500**	0.6000	0.0833	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	7	3	0	1	1.0000	0.6000	0.0000	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	8	2	1	0	0.5000	0.3000	0.5000	0.6333	0.1856	0.6610	-1.1099	1.4812	0.0123	0.7192
Placebo	12	1	1	0	0.0000	0.0000								
	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Drug	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Drug	4	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	6	4	0	1	1.0000	0.8000	0.0000	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	9	3	1	0	0.6667	0.5333	0.1667	0.2167	-0.4642	0.7405	-1.9156	0.9871	0.0683	0.8631
Drug	10	2	1	0	0.5000	0.2667	0.5000	0.7167	0.2790	0.6405	-0.9764	1.5343	0.0097	0.6861
Drug	13	1	1	0	0.0000	0.0000								

Log-Log SE based on Delta Method

$$SE = \left[ln\left[-ln(\widehat{S}(t))\right]\right]$$

$$= \sqrt{\frac{1}{\left[ln(\widehat{S}(t))\right]^2} \sum \frac{d_i}{n_i(n_i - d_i)}}$$

$$= exp(-exp(ln(-ln(\widehat{S}(t)))) \pm 1.96 \times SE(ln(-ln(\widehat{S}(t))))$$

Kaplan-Meier by group

Table 2. Kaplan-Meier survival function by treatment group

Group	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Placebo	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Placebo	1	5	1	0	0.8000	**0.8000**	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Placebo	2	4	1	0	**0.7500**	0.6000	0.0833	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	7	3	0	1	1.0000	0.6000	0.0000	0.1333	-0.6717	0.7148	-2.0727	0.7293	0.1257	0.8818
Placebo	8	2	1	0	0.5000	0.3000	0.5000	0.6333	0.1856	0.6610	-1.1099	1.4812	0.0123	0.7192
Placebo	12	1	1	0	0.0000	0.0000								
	time	n	d	censor	(n-d)/n	S	d/(n(n-d))	sum	ln(-ln(S))	lnln(SE)	lnln(L95%)	lnln(U95%)	L95%CI	U95%CI
Drug	0	5	0	0	1.0000	1.0000	0.0000	0.0000						
Drug	4	5	1	0	0.8000	0.8000	0.0500	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	6	4	0	1	1.0000	0.8000	0.0000	0.0500	-1.4999	1.0021	-3.4640	0.4641	0.2038	0.9692
Drug	9	3	1	0	0.6667	0.5333	0.1667	0.2167	-0.4642	0.7405	-1.9156	0.9871	0.0683	0.8631
Drug	10	2	1	0	0.5000	0.2667	0.5000	0.7167	0.2790	0.6405	-0.9764	1.5343	0.0097	0.6861
Drug	13	1	1	0	0.0000	0.0000								

Log-Log SE based on Delta Method

$$SE = \left[ln\left[-ln(\widehat{S}(t))\right]\right]$$

$$= \sqrt{\frac{1}{\left[ln(\widehat{S}(t))\right]^2} \sum \frac{d_i}{n_i(n_i - d_i)}}$$

$$= exp(-exp(ln(-ln(\widehat{S}(t))))) \pm 1.96 \times SE(ln(-ln(\widehat{S}(t))))$$

Table 3. Long-rank test statistic by treatment group.

	Placebo			Drug					Log-rank estimation		
time	n_{1i}	d_{1i}	c_{1i}	n_{0i}	d_{0i}	c_{0i}	n_i	d_i	e_{1i}	v_{1i}	$d_{1i} - e_{1i}$
0	5	0	0	5	0	0	10	0	0	0	0
1	5	1	0	5	0	0	10	1	0.500	0.250	0.500
2	4	1	0	5	0	0	9	1	0.444	0.247	0.556
4	3	0	0	5	1	0	8	1	0.375	0.234	-0.375
6	3	0	0	4	0	1	7	0	0.000	0.000	0.000
7	3	0	1	3	0	0	6	0	0.000	0.000	0.000
8	2	1	0	3	0	0	5	1	0.400	0.240	0.600
9	1	0	0	3	1	0	4	1	0.250	0.188	-0.250
10	1	0	0	2	1	0	3	1	0.333	0.222	-0.333
12	1	1	0	1	0	0	2	1	0.500	0.250	0.500
13	0	0	0	1	1	0	1	1	0.000	0.000	0.000
										1.63	1.20

Log-rank test

$$\widehat{\mathbf{e}}_{\mathbf{1i}} = \frac{n_{1i} d_i}{n_i} \; ; \mathbf{U} = \sum_{i=1}^{r} \left(d_{1i} - \frac{\mathbf{n_{1i} d_i}}{\mathbf{n_i}} \right) = 1.20$$

Table 3. Long-rank test statistic by treatment group.

	Placebo			Drug					Log-rank estimation		
time	n_{1i}	d_{1i}	c_{1i}	n_{0i}	d_{0i}	c_{0i}	n_i	d_i	e_{1i}	v_{1i}	$d_{1i} - e_{1i}$
0	5	0	0	5	0	0	10	0	0	0	0
1	5	1	0	5	0	0	10	1	0.500	0.250	0.500
2	4	1	0	5	0	0	9	1	0.444	0.247	0.556
4	3	0	0	5	1	0	8	1	0.375	0.234	-0.375
6	3	0	0	4	0	1	7	0	0.000	0.000	0.000
7	3	0	1	3	0	0	6	0	0.000	0.000	0.000
8	2	1	0	3	0	0	5	1	0.400	0.240	0.600
9	1	0	0	3	1	0	4	1	0.250	0.188	-0.250
10	1	0	0	2	1	0	3	1	0.333	0.222	-0.333
12	1	1	0	1	0	0	2	1	0.500	0.250	0.500
13	0	0	0	1	1	0	1	1	0.000	0.000	0.000
										1.63	1.20

Log-rank test

$$\widehat{\mathbf{e}}_{\mathbf{1i}} = \frac{n_{1i} d_i}{n_i} \; ; \mathbf{U} = \sum_{i=1}^{r} \left(d_{1i} - \frac{\mathbf{n_{1i} d_i}}{\mathbf{n_i}} \right) = 1.20$$

$$\widehat{\mathrm{Var}}(\mathbf{U}) = \sum_{i=1}^{r} \frac{n_{1i} n_{0i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = 1.63$$

Inference

Table 3. Long-rank test statistic by treatment group.

	Placebo			Drug					Log-rank estimation		
time	n_{1i}	d_{1i}	c_{1i}	n_{0i}	d_{0i}	c_{0i}	n_i	d_i	e_{1i}	v_{1i}	$d_{1i} - e_{1i}$
0	5	0	0	5	0	0	10	0	0	0	0
1	5	1	0	5	0	0	10	1	0.500	0.250	0.500
2	4	1	0	5	0	0	9	1	0.444	0.247	0.556
4	3	0	0	5	1	0	8	1	0.375	0.234	-0.375
6	3	0	0	4	0	1	7	0	0.000	0.000	0.000
7	3	0	1	3	0	0	6	0	0.000	0.000	0.000
8	2	1	0	3	0	0	5	1	0.400	0.240	0.600
9	1	0	0	3	1	0	4	1	0.250	0.188	-0.250
10	1	0	0	2	1	0	3	1	0.333	0.222	-0.333
12	1	1	0	1	0	0	2	1	0.500	0.250	0.500
13	0	0	0	1	1	0	1	1	0.000	0.000	0.000
										1.63	1.20

Log-rank test

$$\widehat{\mathbf{e}}_{\mathbf{1i}} = \frac{n_{1i} d_i}{n_i} \text{ ; } \mathbf{U} = \sum_{i=1}^{r} \left(d_{1i} - \frac{\mathbf{n_{1i} d_i}}{\mathbf{n_i}} \right) = 1.20$$

$$\widehat{\text{Var}}(\mathbf{U}) = \sum_{i=1}^{r} \frac{n_{1i} n_{0i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = 1.63$$

$$\text{Log-rank test (Q)} = \frac{U^2}{\widehat{\text{Var}}(U)} = \frac{1.43}{1.63} = 0.88 \text{ Chi-square (1df), p-value=}\mathbf{0.348}$$

Inference

Table 3. Long-rank test statistic by treatment group.

	Placebo			Drug					Log-rank estimation		
time	n_{1i}	d_{1i}	c_{1i}	n_{0i}	d_{0i}	c_{0i}	n_i	d_i	e_{1i}	v_{1i}	$d_{1i} - e_{1i}$
0	5	0	0	5	0	0	10	0	0	0	0
1	5	1	0	5	0	0	10	1	0.500	0.250	0.500
2	4	1	0	5	0	0	9	1	0.444	0.247	0.556
4	3	0	0	5	1	0	8	1	0.375	0.234	-0.375
6	3	0	0	4	0	1	7	0	0.000	0.000	0.000
7	3	0	1	3	0	0	6	0	0.000	0.000	0.000
8	2	1	0	3	0	0	5	1	0.400	0.240	0.600
9	1	0	0	3	1	0	4	1	0.250	0.188	-0.250
10	1	0	0	2	1	0	3	1	0.333	0.222	-0.333
12	1	1	0	1	0	0	2	1	0.500	0.250	0.500
13	0	0	0	1	1	0	1	1	0.000	0.000	0.000
										1.63	1.20

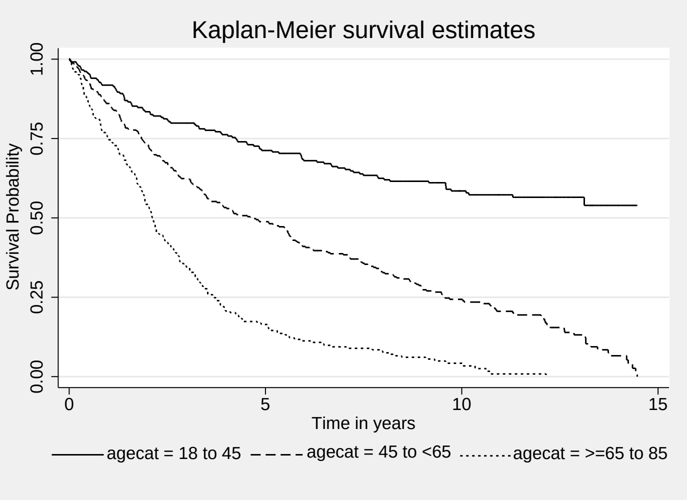
Log-rank test

$$\widehat{\mathbf{e}}_{\mathbf{1i}} = \frac{n_{1i} d_i}{n_i} \; ; \mathbf{U} = \sum_{i=1}^{r} \left(d_{1i} - \frac{\mathbf{n_{1i} d_i}}{\mathbf{n_i}} \right) = 1.20$$

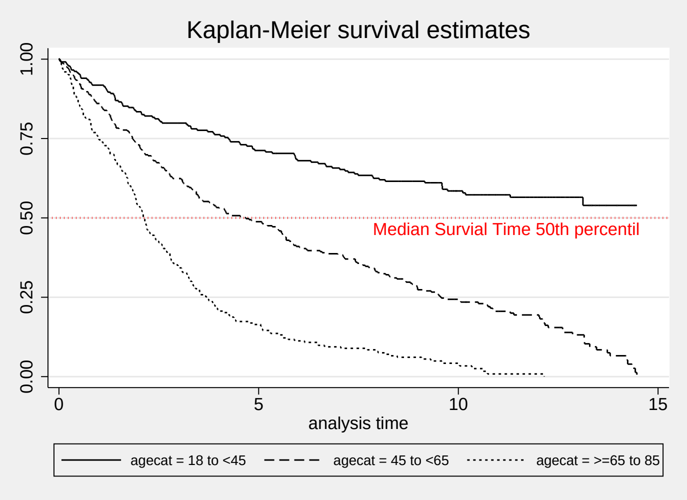
$$\widehat{\text{Var}}(\mathbf{U}) = \sum_{i=1}^{r} \frac{n_{1i} n_{0i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)} = 1.63$$

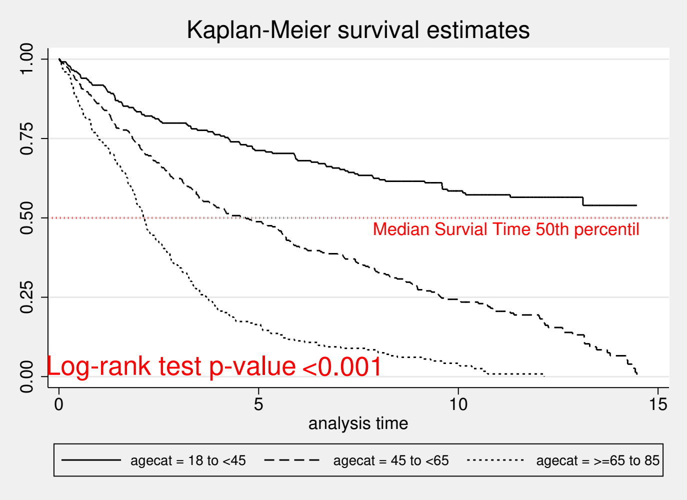
$$\text{Log-rank test (Q)} = \frac{U^2}{\widehat{\text{Var}}(U)} = \frac{1.43}{1.63} = 0.88 \text{ Chi-square (1df), p-value=}\mathbf{0.348}$$

Kaplan-Meier



Kaplan-Meier




Kaplan-Meier survival estimates
1.00
0.75
0.50
0.25
0.00
0
5
10
15
analysis time
Median Survial Time 50th percentil
Log-rank test p-value <0.001
agecat = 18 to <45
agecat = 45 to <65
agecat = >=65 to 85

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models (HM)

- There are additive

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models (HM)

- There are additive and multiplicative

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models (HM)

- There are additive and multiplicative (proportional) hazard models

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models (HM)

- There are additive and multiplicative (proportional) hazard models
- Hazard models are the basis of Parametric and Semi-parametric multivariate time-to-event analysis

Hazard Models

Why we need them?

- Non-parametric methods cannot control for confounding (univariate or just bivariate)
- We need multivariate regression modelling techniques to adjust for confounding

Hazard Models (HM)

- There are additive and multiplicative (proportional) hazard models
- Hazard models are the basis of Parametric and Semi-parametric multivariate time-to-event analysis

Hazard Models (HM)

The parametrization of the HM explains all:

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} =$$

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)}$$

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times$$

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{j=1}^{k} \beta \mathbf{x_j})$$

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_\mathbf{0} + \sum_{\mathbf{j=1}}^{\mathbf{k}} \beta \mathbf{x_j})$$

Functional form of the baseline hazard

- Assumptions about the **functional form** of the $h_0(t)$ is what determines whether the model is parametric or semiparametric

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{j=1}^{k} \beta \mathbf{x_j})$$

Functional form of the baseline hazard

- Assumptions about the **functional form** of the $h_0(t)$ is what determines whether the model is parametric or semiparametric
- The **exponential** of the function of the covariates is needed to prevent a negative baseline hazard ($h_0(t)$)

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{\mathbf{j=1}}^{\mathbf{k}} \beta \mathbf{x_j})$$

Functional form of the baseline hazard

- Assumptions about the **functional form** of the $h_0(t)$ is what determines whether the model is parametric or semiparametric
- The **exponential** of the function of the covariates is needed to prevent a negative baseline hazard ($h_0(t)$)

Hazard Models (HM)

The parametrization of the HM explains all:

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} =$$

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)}$$

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times$$

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{j=1}^{k} \beta \mathbf{x_j})$$

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{j=1}^{k} \beta \mathbf{x_j})$$

Functional form of the baseline hazard

- Flexible parametric models allow smooth functions for the baseline hazard relaxing the assumptions about the **parametric functional form** of the $h_0(t)$

Hazard Models

Hazard Models (HM)

The parametrization of the HM explains all:

$$\mathbf{h_j(t|x)} = \mathbf{h_0(t)} \times \mathbf{exp}(\beta_0 + \sum_{j=1}^{k} \beta \mathbf{x_j})$$

Functional form of the baseline hazard

- Flexible parametric models allow smooth functions for the baseline hazard relaxing the assumptions about the **parametric functional form** of the $h_0(t)$

Hazard Models

Proportionality assumption

- The proportional hazards assumption states that the **model covariates** are multiplicatively related to the **baseline hazard** (assumption that may or may not be true)

Hazard Models

Proportionality assumption

- The proportional hazards assumption states that the **model covariates** are multiplicatively related to the **baseline hazard** (assumption that may or may not be true)
- Allowing for the interaction between the baseline hazard $h_0(t)$ and some of the predictors (βx) relax the **proportionality of hazard** assumption

Hazard Models

Proportionality assumption

- The proportional hazards assumption states that the **model covariates** are multiplicatively related to the **baseline hazard** (assumption that may or may not be true)
- Allowing for the interaction between the baseline hazard $h_0(t)$ and some of the predictors (βx) relax the **proportionality of hazard** assumption

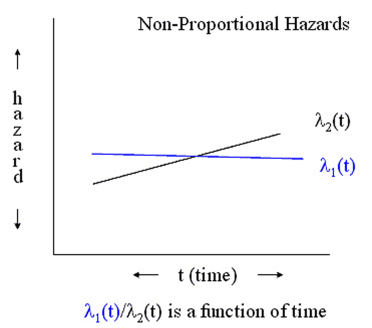
Proportionality assumption

Hazard Models

Proportionality assumption

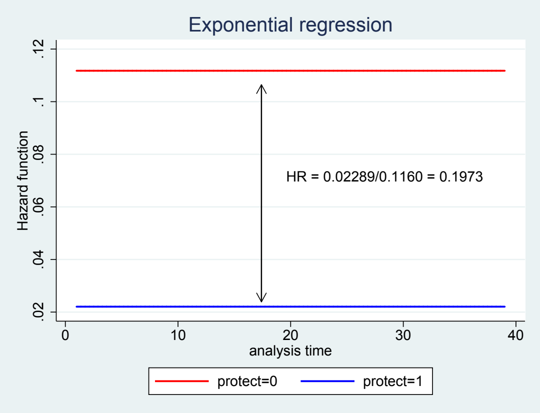
- The proportional hazards assumption states that the **model covariates** are multiplicatively related to the **baseline hazard** (assumption that may or may not be true)
- Allowing for the interaction between the baseline hazard $h_0(t)$ and some of the predictors (βx) relax the **proportionality of hazard** assumption

Proportionality assumption



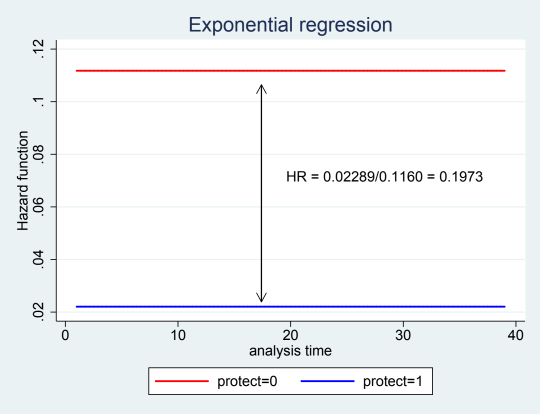
Proportionality of Hazards

Exponential model with constant Hazard



Proportionality of Hazards

Exponential model with constant Hazard



Exponential Models

```
.use http://www.stata.com/courses/nc631-13/lec4/hip2, clear
. streg protect, dist(exp) hr
         failure _d:  fracture
   analysis time _t:  time1
                 id:  id
Exponential regression -- log relative-hazard form
No. of subjects =           48                 Number of obs    =        106
No. of failures =           31
Time at risk    =          714
                                               LR chi2(1)       =      19.68
Log likelihood  =   -50.227658                 Prob > chi2      =     0.0000
------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     protect |   .1973684   .0727765    -4.40   0.000     .0958099    .4065789
       _cons |   .1117647   .0256406    -9.55   0.000     .0712895      .17522
------------------------------------------------------------------------------
```

Parametric Survival Distributions: Pintilie M. (2007)

Table 2.2 Parametric models for time-to-event data.

Distribution	$f(t)$	$F(t)$	$S(t)$	$h(t)$
Exponential	$\lambda e^{-\lambda t}$	$1-e^{-\lambda t}$	$e^{-\lambda t}$	λ
Weibull	$\lambda\theta t^{\theta-1}\exp\left\{-\lambda t^{\theta}\right\}$	$1-\exp\left\{-\lambda t^{\theta}\right\}$	$\exp\left\{-\lambda t^{\theta}\right\}$	$\lambda\theta t^{\theta-1}$
Gamma	$\lambda^{\theta}t^{\theta-1}e^{-\lambda t}$	$I\Gamma(\theta,\lambda t)^{\dagger}$	$1-I\Gamma(\theta,\lambda t)$	$f(t)/S(t)$
Generalized gamma	$\dfrac{\alpha\lambda^{\theta}t^{\alpha\theta-1}\exp\left\{-\lambda t^{\alpha}\right\}}{\Gamma(\theta)}$	$I\Gamma(\theta,\lambda t^{\alpha})^{\dagger}$	$1-I\Gamma(\theta,\lambda t^{\alpha})$	$f(t)/S(t)$
Lognormal	$\dfrac{\exp\left\{-\frac{1}{2}\left(\frac{\log t-\mu}{\sigma}\right)^{2}\right\}}{\sigma t\sqrt{2\pi}}$	$\Phi\left(\dfrac{\log t-\mu}{\sigma}\right)^{\ddagger}$	$1-\Phi\left(\dfrac{\log t-\mu}{\sigma}\right)$	$f(t)/S(t)$
Log-logistic	$\dfrac{\lambda\theta t^{\theta-1}}{(1+\lambda t^{\theta})^{2}}$	$\dfrac{\lambda t^{\theta}}{1+\lambda t^{\theta}}$	$\dfrac{1}{1+\lambda t^{\theta}}$	$\dfrac{\lambda\theta t^{\theta-1}}{1+\lambda t^{\theta}}$

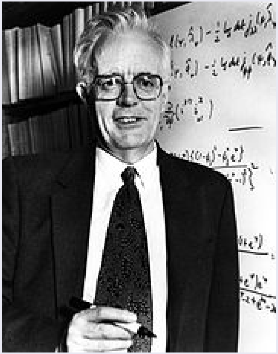
† $I\Gamma$ is the incomplete gamma function: $I\Gamma(\theta,\lambda t)=\int_0^t \lambda^{\theta}t^{\theta-1}e^{-\lambda t}dt/\Gamma(\theta)$.
‡ Φ is the cdf of the standard normal distribution.

Shiny application

http://watzilei.com/shiny

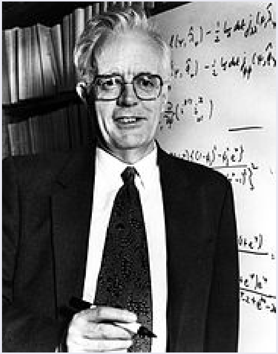
Semi-parametric Cox Proportional Hazard Model

Sir David Cox



Semi-parametric Cox Proportional Hazard Model

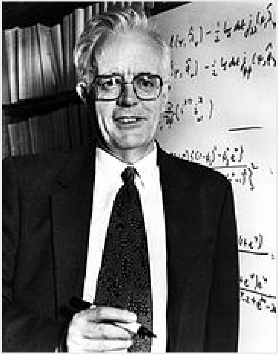
Sir David Cox



Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

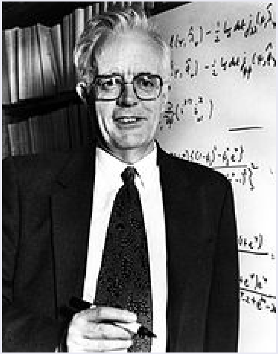


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) =$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

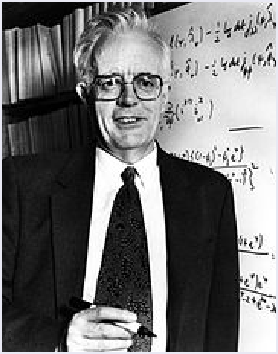


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t})$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

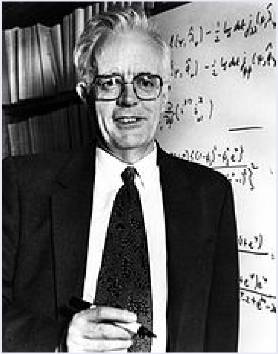


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)]$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

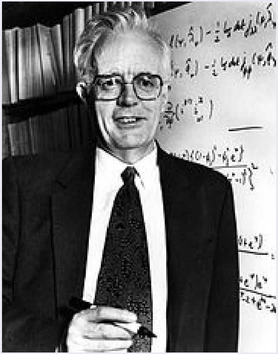


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

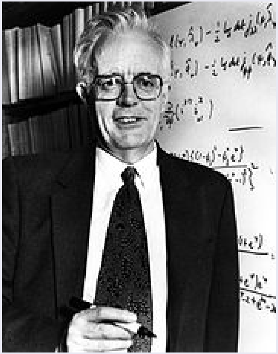


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times \mathbf{exp}(\beta \mathbf{x_j})$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox

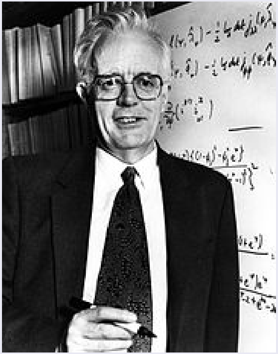


Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times \mathbf{exp}(\beta \mathbf{x_j})$$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox



Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

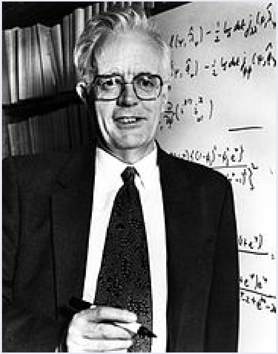
$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times \mathbf{exp}(\beta \mathbf{x_j})$$

Semiparametric Cox Model

- Assumptions about the functional form of the $h_0(t)$ is **unspecified**

Semi-parametric Cox Proportional Hazard Model

Sir David Cox



Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

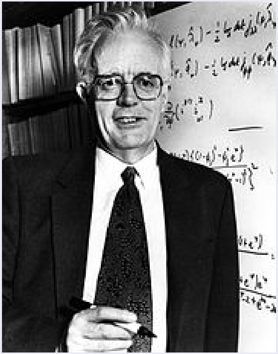
$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times \mathbf{exp}(\beta \mathbf{x_j})$$

Semiparametric Cox Model

- Assumptions about the functional form of the $h_0(t)$ is **unspecified**
- The constant of the model cannot be estimated and is absorbed into the baseline hazard $h_0(t)$

Semi-parametric Cox Proportional Hazard Model

Sir David Cox



Sir David Cox has made pioneering and important contributions to numerous areas of statistics and applied probability, of which the best known is perhaps the semi-parametric proportional hazards model

$$\mathbf{h}(\mathbf{t}|\mathbf{x_j}) = [\mathbf{h_0}(\mathbf{t}) \times exp(\beta_0)] \times \mathbf{exp}(\beta \mathbf{x_j})$$

Semiparametric Cox Model

- Assumptions about the functional form of the $h_0(t)$ is **unspecified**
- The constant of the model cannot be estimated and is absorbed into the baseline hazard $h_0(t)$

Cox Model in Stata

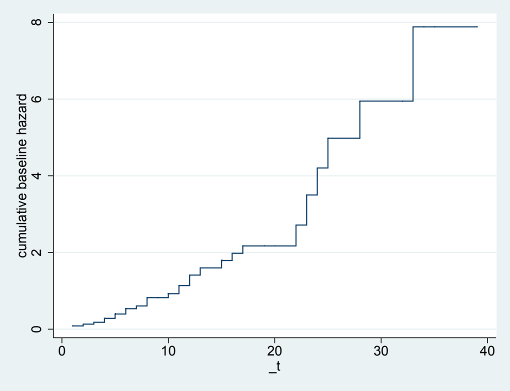
```
. use http://www.stata.com/courses/nc631-13/lec4/hip2, clear
(hip fracture study)
. stcox protect, nohr
failure _d: fracture
analysis time _t: time1
id: id
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     protect |  -2.047599   .4404029    -4.65   0.000    -2.910773   -1.184426
------------------------------------------------------------------------------

Hazard Ratio for wearing the protect device = ? ==> exp(-2.05) = 0.129

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     protect |   .1290443   .0568315    -4.65   0.000     .0544336    .3059218
------------------------------------------------------------------------------
```

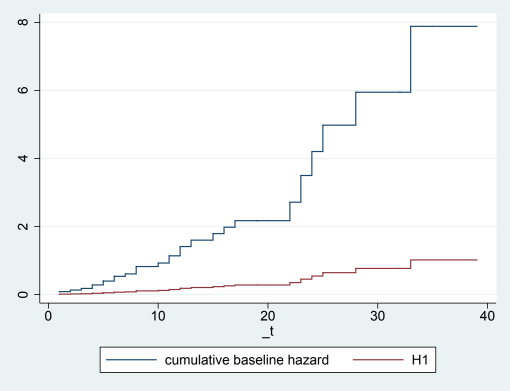
Cumulative Hazard in Stata

- predict H0, basechazard
- twoway line H0 _*t*, connect(J) sort



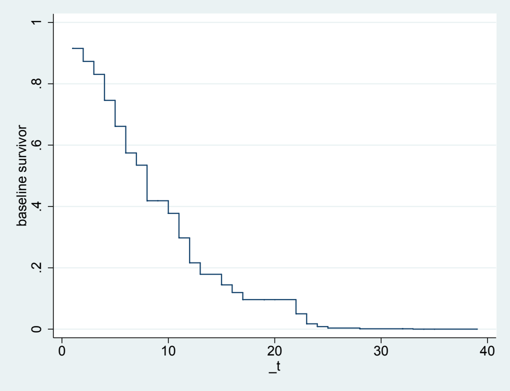
Cumulative Hazard in Stata

- generate H1 = H0*exp(_b[protect])
- twoway line H0 H1 _t, sort connect(J J)
- stcurve, cumhaz at1(protect = 0) at2(protect=1)



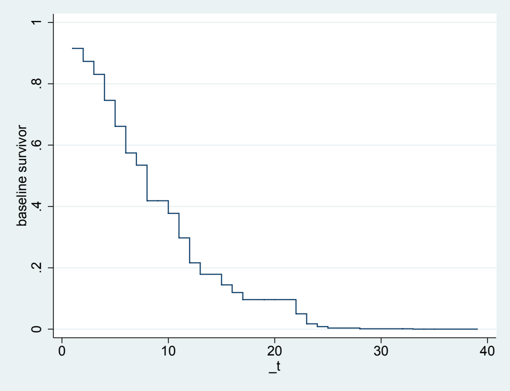
Baseline Survival in Stata

- predict S0, basesurv
- twoway line S0 _*t*, connect(J) sort



Baseline Survival in Stata

- predict S0, basesurv
- twoway line S0 _t, connect(J) sort

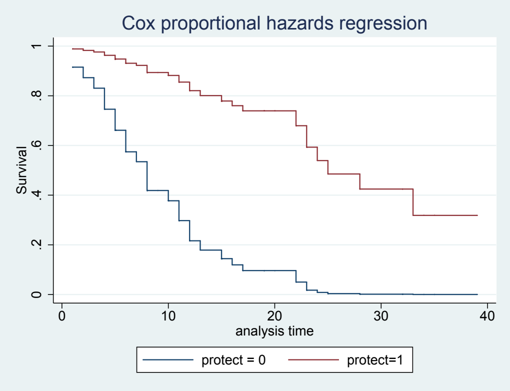


Survival by groups in Stata

- generate S1 = S0*exp(_*b*[protect])
- twoway line S0 S1 _*t*, sort connect(J J)
- stcurve, surv at1(protect = 0) at2(protect=1)

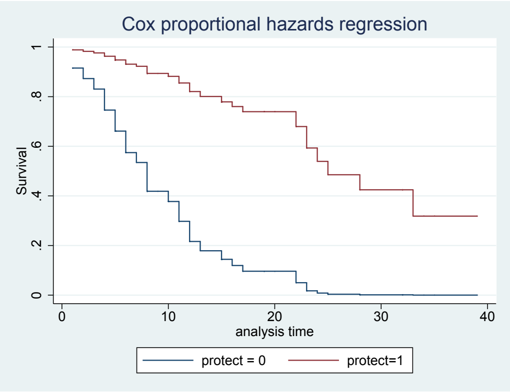
Survival by groups in Stata

- generate S1 = S0*exp(_b[protect])
- twoway line S0 S1 _t, sort connect(J J)
- stcurve, surv at1(protect = 0) at2(protect=1)



Survival by groups in Stata

- generate S1 = S0*exp(_b[protect])
- twoway line S0 S1 _t, sort connect(J J)
- stcurve, surv at1(protect = 0) at2(protect=1)



References

- Klein, J. P., and M. L. Moeschberger. 2003. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York: Springer.
- Mantel, N., and W. Haenszel. 1959. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22: 719-748.
- Nelson, W. 1972. Theory and applications of hazard plotting for censored failure data. Technometrics 14: 945965.

References

References

- Peto, R., and J. Peto. 1972. Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, Series A 135: 185207.
- Prentice, R. L. 1978. Linear rank tests with right censored data. Biometrika 65: 167179.
- Tarone, R. E., and J. H. Ware. 1977. On distribution-free tests for equality of survival distributions. Biometrika 64: 156-1
- Aalen, O. 1978. Nonparametric inference for a family of counting processes. Annals of Statistics 6: 701726.

References

- Breslow, N. E. 1970. A generalized KruskalWallis test for comparing K samples subject to unequal patterns of censorship. Biometrika 57: 579594.
- Colosimo, E., F. Ferreira, M. Oliveira, and C. Sousa. 2002. Empirical comparisons between Kaplan-Meier and
- Nelson-Aalen survival function estimators. Journal of Statistical Computation and Simulation 72: 299308.

References

- Gehan, E. A. 1965. A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. Biometrika 52: 203223.
- Harrington, D. P., and T. R. Fleming. 1982. A class of rank test procedures for censored survival data. Biometrika 69: 553566.
- Hosmer, D. W., Jr., S. A. Lemeshow, and S. May. 2008. Applied Survival Analysis: Regression Modeling of Time to Event Data. 2nd ed. New York: Wiley.
- Kaplan, E. L., and P. Meier. 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53: 457481.

References

- Cox, D. R. 1972. Regression models and life-tables. Journal of the Royal Statistical Society, Series B 34.
- Cox, D. R., and E. J. Snell. 1968. A general definition of residuals. Journal of the Royal Statistical Society, Series B 30: 248-275.
- Efron, B. 1977. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association 72: 557-565.
- Garrett, J. M. 1997. gr23: Graphical assessmenat of the Cox model proportional hazards assumption. Stata Technical Bulletin 35: 9-14. Reprinted in Stata Technical Bulletin Reprints, vol. 6, pp. 38-44. College Station, TX: Stata Press.

References

- Grambsch, P. M., and T. M. Therneau. 1994. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika 81: 515-526.
- Hess, K. R. 1995. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. Statistics in Medicine 14: 1707-1723.
- Royston, P., and P. C. Lambert. 2011. Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model. College Station, TX: Stata Press.

Final overview

https://www.mailman.columbia.edu/research/population-health-methods/time-event-data-analysis

Thank you!

Muchas Gracias!