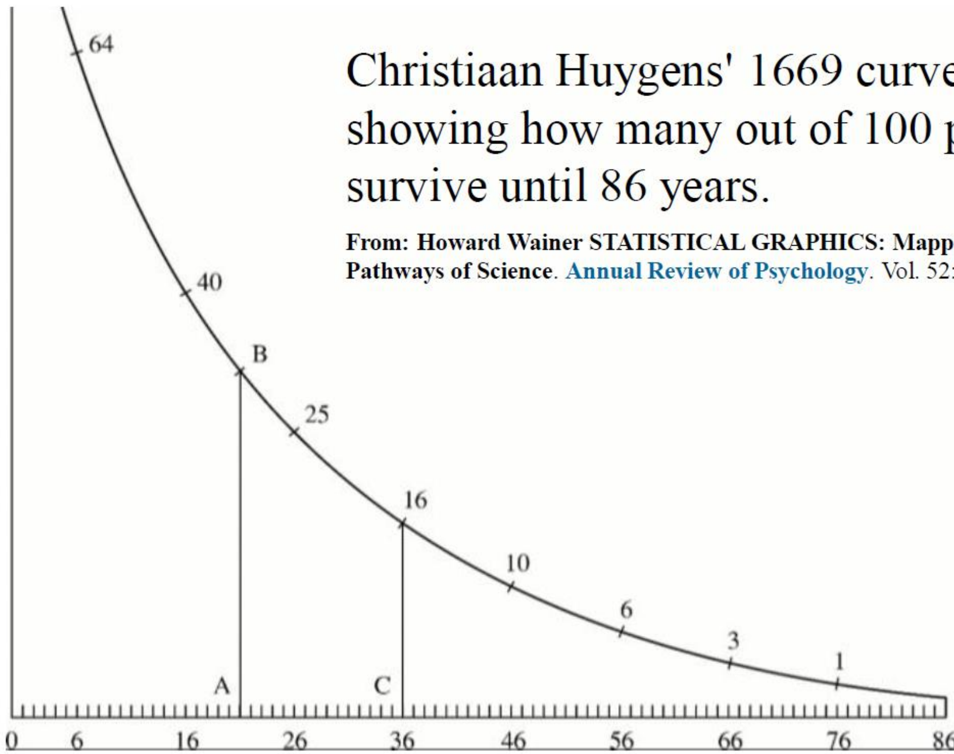
Survival Analysis

STATS – Modelação Estatística

PhD Programme in Health Data Science

Cristina Costa Santos & Andreia Teixeira

One of the first exemples of survival analysis 1669



Christiaan Huygens' 1669 curve showing how many out of 100 people survive until 86 years.

From: Howard Wainer STATISTICAL GRAPHICS: Mapping the Pathways of Science. Annual Review of Psychology. Vol. 52: 305-335.

What is survival analysis?

HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

- Statistical methods for analyzing **longitudinal time to event data**.

- **Events** may include: death, injury, onset of illness, recovery from illness or transition above or below the clinical threshold of a meaningful continuous variable.

Objectives of survival analysis

- **Estimate time-to-event for a group of individuals**, such as time until death for a group of cancer patients.

- **To compare time-to-event between two or more groups**, such as treated vs placebo cancer patients in a randomized controlled trial.

- **To assess the relationship of co-variables to time-to-event**, such as: does age, a certain biomarker, or tumor size influence survival time of cancer patients?

Censoring

- **The patients who did not have the event are considered censored.**

  - **We know that they survived a specific amount of time, but do not know the exact time of the event.**
  - **We believe that the event would have happened if we observed them long enough.**

- **These patients provide some information, but not complete information.**

Censoring

- **How could we account for censoring?**

  - **Ignore it and say event ocurred at time of censoring**
    - Incorrect because this is almost certainly not true
  - **Remove patient from analysis**
    - Potential bias and loss of power
  - **Survival analysis**

- **Our objective is to estimate the survival distribution of patients in the presence of censoring.**

→ **Censoring is taken into account in the analysis.**

Why use survival analysis?

- **Why not compare mean time-to-event between groups using a t-test or linear regression?**

- **Why not compare proportion of events in groups using risk/odds ratios or logistic regression?**

Why use survival analysis?

- **Why not compare mean time-to-event between groups using a t-test or linear regression?**
  - Ignores censoring

- **Why not compare proportion of events in groups using risk/odds ratios or logistic regression?**
  - Ignores time

Why use survival analysis?

- In survival analysis the time until the occurrence of a well-defined event is recorded (time-to-event data).
  - Survival time

- If everyone had an event, some of the methods we have already learned could be applied.

- Often, not everyone has event - censoring
  - Loss to follow-up
  - End of study

Survival analysis: Terms

- **Time-to-event:** the time from entry into a study until a subject has a particular outcome

- **Censoring:** subjects are said to be censored if they are lost to follow up or drop out of the study, or if they die of unrelated causes or the study ends before they have the event of interest. They are counted as alive or disease-free for the time they were enrolled in the study.

Data Structure: survival analysis

Two-variable outcome:

- **Time variable (T):** t_i = time at last event-free observation or time at event
- **Censoring variable (C):** c_i = 1 if had the event; c_i = 0 no event by time t_i

Introduction to survival distributions

- T_i the event time for an individual, is a random variable having a probability distribution.

- Different models for survival data are distinguished by different choice of distribution for T_i.

- The probability (density) of the event time occurring at exactly time t:

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t)}{\Delta t}$$

- The cumulative probability (density) of the event time ocurring before time t:

$$F(t) = P(T \leq t) = \int_{-\infty}^{t} f(t)dt$$

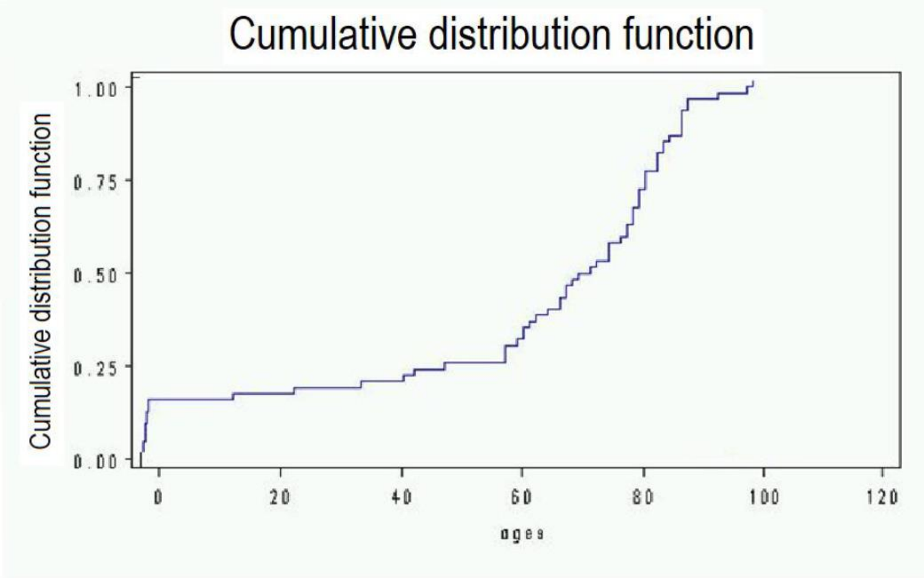
Survival function: 1-F(t)

The goal of survival analysis is to estimate and compare survival experiences of diferent groups.

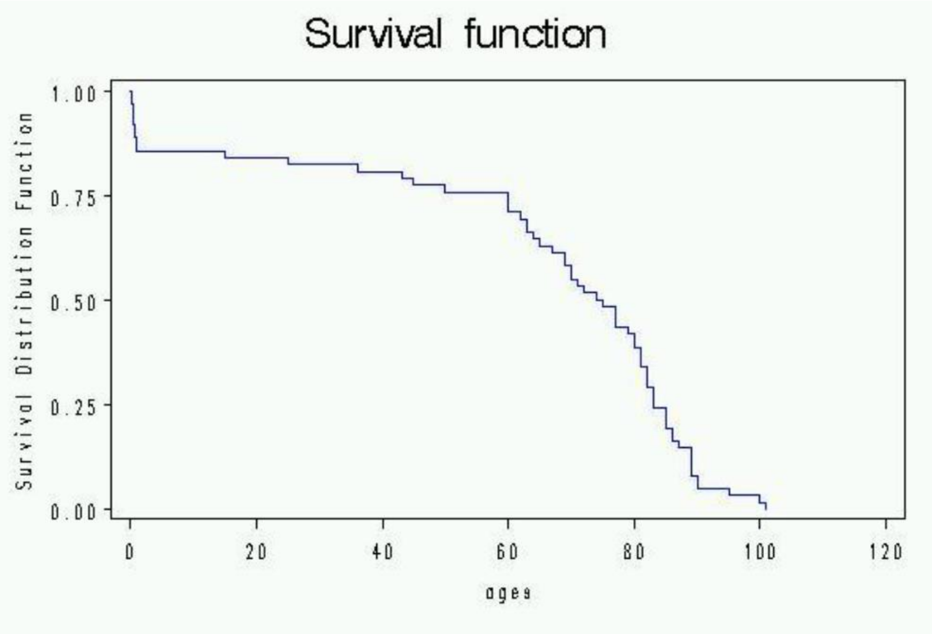
Survival experience is described by the cumulative survival function:

$$S(t) = 1 - P(T \leq t) = 1 - F(t)$$

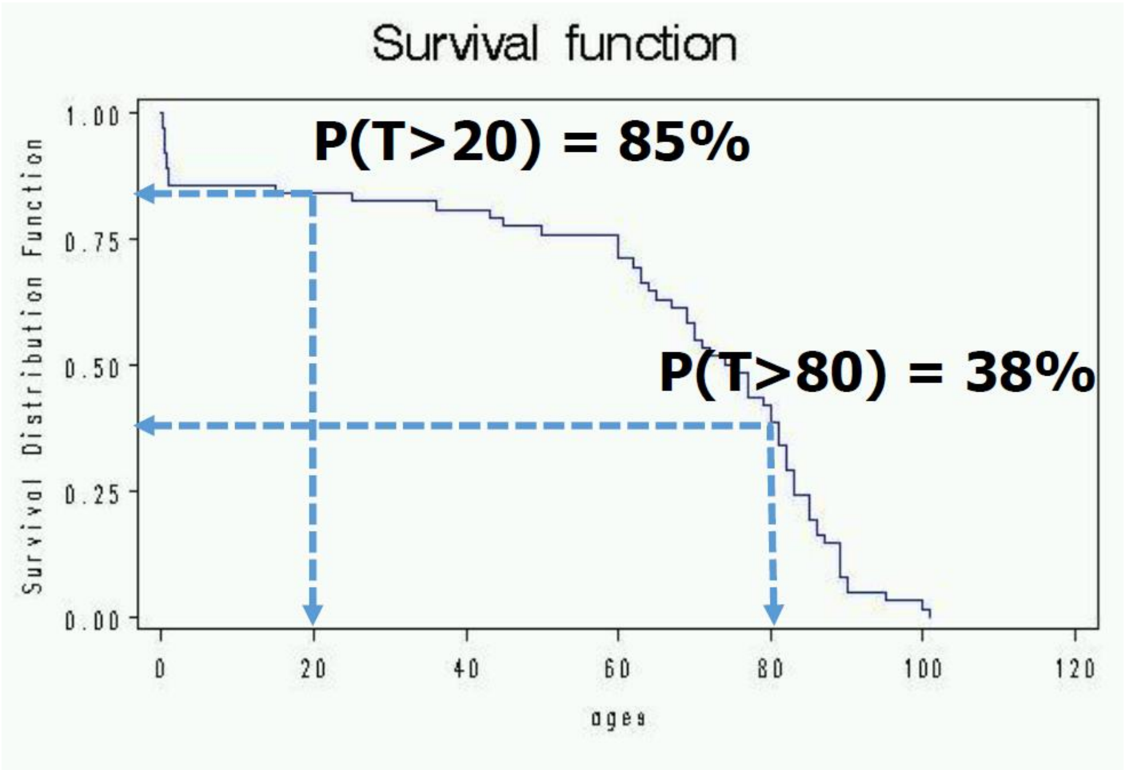
Example: If t = 100 years, S(t = 100) = probability of surviving beyond 100 years.


Cumulative distribution function
Cumulative distribution function
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
100
120
ages

F(t)


Survival function
Survival Distribution Function
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
100
120
ages

S(t)


Survival function
P(T>20) = 85%
P(T>80) = 38%
Survival Distribution Function
1.00
0.75
0.50
0.25
0.00
0
20
40
60
80
100
120
ages

Hazard Function: new concept

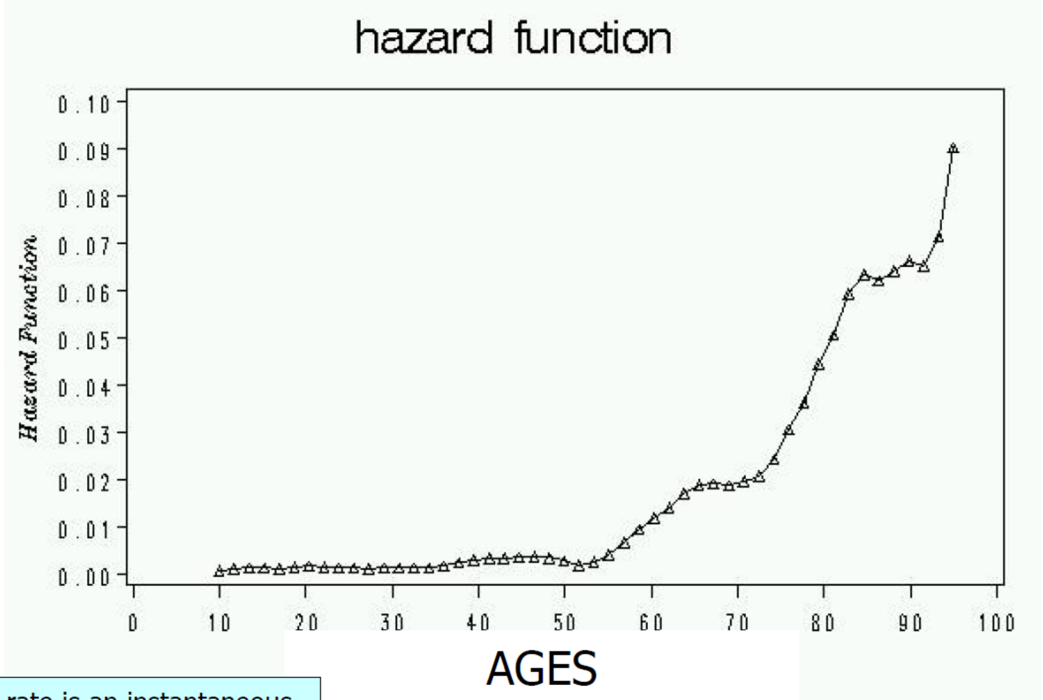
- The probability that if you survive to t, you will succumb to the event in the next instant:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t \,|T \ \ge t)}{\Delta t}$$

$$h(t)dt = P(t \ \le T < t + dt \,|T \ \ge t) = \frac{P(t \ \le T < t + dt \ \wedge T \ \ge t)}{P(T \ge t)} = \frac{P(t \le T < t + dt)}{P(T \ge t)} = \frac{f(t)dt}{S(t)} = \frac{-S'(t)dt}{S(t)}$$

$$h(t) = \frac{-S'(t)}{S(t)}$$

Hazard Function: new concept



Hazard rate is an instantaneous incidence rate.

Hazard vs Density example

- When you are born, you have a certain probability of dying at any age; that's the probability density

(*Example:* a woman born today has, say, a 1% chance of dying at 80 years).

- However, as you survive for awhile, your probabilities keep changing; that's the hazard rate (think: conditional probability)

(*Example:* a woman who is 79 today has, say, a 5% chance of dying at 80 years).

Practical example
Basics of Survival Analysis

In **R** (as with other packages) we require the following **two variables when dealing with survival time data**:

- A **continuous time variable** that measures the time until either the event or the individuals withdrawal (censoring).

- **A categorical variable that acts as an indicator for whether the subject experienced the event of interest or whether they did not and were censored.**

Kaplan-Meier Analysis Example

Rather than categorising, we can estimate the survival function directly from the continuous survival times.

File: data2.csv

Kaplan-Meier Analysis Example

The **variables** included in the data are:

- id: identification of the patient
- fu: survival time in months
- event: event observed [0 = censored; 1 = death]
- eventHD: event observed [0 = censored; 1 = transfer to haemodialysis]
- eventTR: event observed [0 = censored; 1 = renal transplant]
- eventCR: event observed considering a competing risks approach [0 = censored; 1 = death; 2 = transfer to haemodialysis; 3 = renal transplant]
- peritonitis: occurence of a peritonites [0 = no; 1 = yes]
- sex [0 = female; 1 = male]
- age: in years
- diab: diabetes [0 = no; 1 = yes]
- first: prior treatment [0 = no; 1 = yes]

Kaplan-Meier Analysis Example

```
#Load the library required for a survival analysis

library(splines)
library(survival)

#Read data

data=read.csv("data2.csv", sep=";", dec=",")
```

Kaplan-Meier Analysis Example

#Read data

```
data=read.csv("data2.csv", sep=";", dec=",")
```

```
> head(data)
  id fu event event_HD event_RT event_CR peritonitis sex age diab first
1  1  7     1        0        0        1           0   1  60    1     0
2  3 22     0        0        1        3           1   1  41    1     0
3  7 13     0        0        1        3           1   0  38    0     1
4  9 42     1        0        0        1           1   0  46    0     0
5 11  5     0        1        0        2           1   0  51    0     0
6 13 13     1        0        0        1           0   0  36    0     1
```

Kaplan-Meier Analysis Example

#Provides the names of variables included in the database

```
> names(data)
 [1] "id"          "fu"          "event"       "event_HD"    "event_RT"    "event_CR"    "peritonitis"
 [8] "sex"         "age"         "diab"        "first"
```

#Provides dimension of the database

```
> dim(data)
[1] 274  11
```

Kaplan-Meier Analysis Example

The first step is to create a survival object, considering the variables follow-up and event. Note that a “+” after the time indicates censoring.

#Create a survival object

```
> Surv(data$fu,data$event)
  [1]   7.00   22.00+  13.00+  42.00    5.00+  13.00   70.00+  51.00   25.00+   5.00+  17.00   42.00+   6.00 
 [14]  10.00+  49.00+  55.00+  18.00+   7.00+   2.00+  22.00   24.00    2.00+   1.00+  10.00   17.00+  35.00 
 [27]  32.00+  30.00+   2.00    5.00    6.00+  78.00+   4.00+  63.00+   6.00+   9.00   22.00+   6.00+  33.00+
 [40]   1.00+   9.00   27.00+   2.00+  27.00   21.00+  14.00+   2.00   53.00+  38.00   62.00+  44.00+  58.00+
 [53]  33.00+  23.00   22.00+  26.00+  11.00+   2.00   28.00+  74.00+  28.00+  13.00   17.00+  30.00+  24.00+
 [66] 109.00+ 105.00+  36.00+  14.00+   7.00+  43.00+   6.57+  37.00   22.00+   1.00   22.00+  84.00+  37.00 
 [79]  14.00   60.00    8.00   35.00   23.00+  21.00+  92.00+  36.00   31.00+   2.00+  28.00+  13.00   12.00 
 [92]   8.00+   7.00+  20.00+  18.00   40.00+  78.00+  61.00+  10.00+ 101.00+   2.00+   9.00+   4.00+  83.00+
[105]   8.00+   9.00+   1.00+  34.00+  33.00+  11.00+  32.00+  31.00+  29.00+   5.00+  21.00+  26.00+  26.00+
[118]  24.00+  24.00+   8.00+  14.00+  19.00+   8.00+  18.00+  16.00+  16.00+  14.00+  15.00+  10.00+  10.00+
[131]  22.00+   8.00+   7.00+   6.00+   5.00+   3.00+   2.00+   7.00   22.00+  13.00+  42.00    5.00+  13.00 
[144]  70.00+  51.00   25.00+   5.00+  17.00   42.00+   6.00   10.00+  49.00+  55.00+  18.00+   7.00+   2.00+
[157]  22.00   24.00    2.00+   1.00+  10.00   17.00+  35.00   32.00+  30.00+   2.00    5.00    6.00+  78.00+
[170]   4.00+  63.00+   6.00+   9.00   22.00+   6.00+  33.00+   1.00+   9.00   27.00+   2.00+  27.00   21.00+
[183]  14.00+   2.00   53.00+  38.00   62.00+  44.00+  58.00+  33.00+  23.00   22.00+  26.00+  11.00+   2.00 
[196]  28.00+  74.00+  28.00+  13.00   17.00+  30.00+  24.00+ 109.00+ 105.00+  36.00+  14.00+   7.00+  43.00+
[209]   6.57+  37.00   22.00+   1.00   22.00+  84.00+  37.00   14.00   60.00    8.00   35.00   23.00+  21.00+
[222]  92.00+  36.00   31.00+   2.00+  28.00+  13.00   12.00    8.00+   7.00+  20.00+  18.00   40.00+  78.00+
[235]  61.00+  10.00+ 101.00+   2.00+   9.00+   4.00+  83.00+   8.00+   9.00+   1.00+  34.00+  33.00+  11.00+
[248]  32.00+  31.00+  29.00+   5.00+  21.00+  26.00+  26.00+  24.00+  24.00+   8.00+  14.00+  19.00+   8.00+
[261]  18.00+  16.00+  16.00+  14.00+  15.00+  10.00+  10.00+  22.00+   8.00+   7.00+   6.00+   5.00+   3.00+
[274]   2.00+
```

Kaplan-Meier Analysis Example

To produce the Kaplan-Meier estimates of the probability of survival over time. The median of survival and the survival in several points of time could be requested.

#Kaplan-Meier survival curve

```
> KM_1 <- survfit(Surv(data$fu,data$event)~1)
> KM_1
Call: survfit(formula = Surv(data$fu, data$event) ~ 1)

      n  events  median 0.95LCL 0.95UCL
    274      62      NA      51      NA
> summary(KM_1)
Call: survfit(formula = Surv(data$fu, data$event) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    274       2    0.993 0.00514        0.983        1.000
    2    266       6    0.970 0.01034        0.950        0.991
    5    242       2    0.962 0.01171        0.940        0.986
    6    232       2    0.954 0.01299        0.929        0.980
    7    220       2    0.945 0.01425        0.918        0.974
    8    210       2    0.936 0.01547        0.906        0.967
    9    198       4    0.917 0.01782        0.883        0.953
   10    190       2    0.908 0.01889        0.871        0.946
   12    176       2    0.897 0.02004        0.859        0.938
   13    174       6    0.866 0.02299        0.823        0.913
   14    166       2    0.856 0.02386        0.811        0.904
   17    150       2    0.845 0.02487        0.797        0.895
   18    144       2    0.833 0.02587        0.784        0.885
   22    128       2    0.820 0.02706        0.769        0.875
   23    114       2    0.806 0.02843        0.752        0.863
   24    110       2    0.791 0.02974        0.735        0.851
   27     94       2    0.774 0.03140        0.715        0.838
   35     62       4    0.724 0.03802        0.653        0.803
   36     58       2    0.699 0.04061        0.624        0.783
   37     54       4    0.647 0.04510        0.565        0.742
   38     50       2    0.621 0.04687        0.536        0.720
   42     46       2    0.594 0.04857        0.506        0.698
   51     36       2    0.561 0.05118        0.470        0.671
   60     28       2    0.521 0.05482        0.424        0.641
```

Kaplan-Meier Analysis Example

To produce the Kaplan-Meier estimates of the probability of survival over time. The median of survival and the survival in several points of time could be requested.

#Kaplan-Meier survival curve

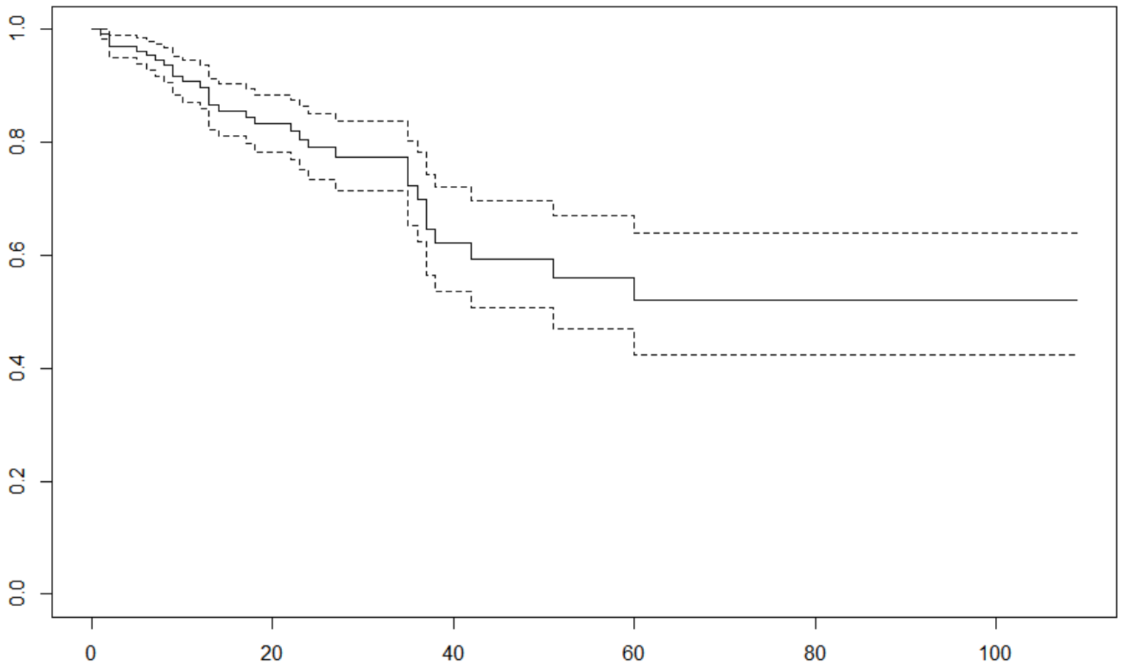
The **median survival** is the smallest time at which the survival probability drops to 0.5 (50%) or below. If the survival curve does not drop to 0.5 or below then the median time cannot be computed.

```
> KM_1 <- survfit(Surv(data$fu,data$event)~1)
> KM_1
Call: survfit(formula = Surv(data$fu, data$event) ~ 1)

      n  events  median 0.95LCL 0.95UCL
    274      62      NA      51      NA
> summary(KM_1)
Call: survfit(formula = Surv(data$fu, data$event) ~ 1)

 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    274       2    0.993 0.00514        0.983        1.000
    2    266       6    0.970 0.01034        0.950        0.991
    5    242       2    0.962 0.01171        0.940        0.986
    6    232       2    0.954 0.01299        0.929        0.980
    7    220       2    0.945 0.01425        0.918        0.974
    8    210       2    0.936 0.01547        0.906        0.967
    9    198       4    0.917 0.01782        0.883        0.953
   10    190       2    0.908 0.01889        0.871        0.946
   12    176       2    0.897 0.02004        0.859        0.938
   13    174       6    0.866 0.02299        0.823        0.913
   14    166       2    0.856 0.02386        0.811        0.904
   17    150       2    0.845 0.02487        0.797        0.895
   18    144       2    0.833 0.02587        0.784        0.885
   22    128       2    0.820 0.02706        0.769        0.875
   23    114       2    0.806 0.02843        0.752        0.863
   24    110       2    0.791 0.02974        0.735        0.851
   27     94       2    0.774 0.03140        0.715        0.838
   35     62       4    0.724 0.03802        0.653        0.803
   36     58       2    0.699 0.04061        0.624        0.783
   37     54       4    0.647 0.04510        0.565        0.742
   38     50       2    0.621 0.04687        0.536        0.720
   42     46       2    0.594 0.04857        0.506        0.698
   51     36       2    0.561 0.05118        0.470        0.671
   60     28       2    0.521 0.05482        0.424        0.641
```

Kaplan-Meier Analysis Example

#Plot survival curve with confidence intervals

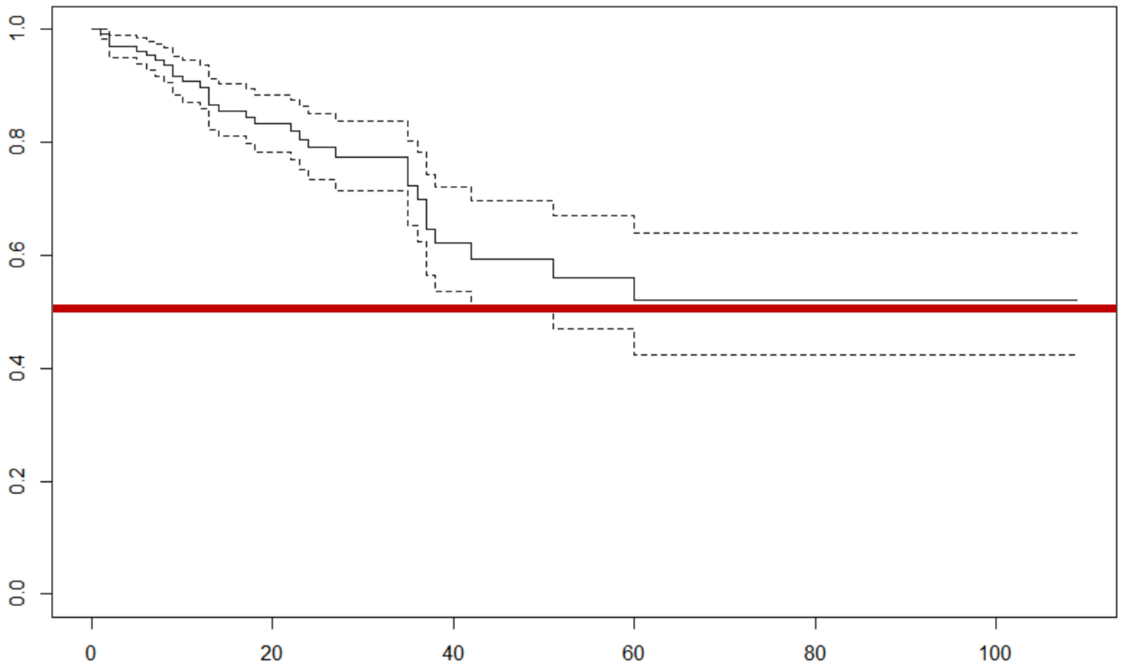
```
> plot(KM_1, mark.time=F)
```



Kaplan-Meier Analysis Example

#Plot survival curve with confidence intervals

```
> plot(KM_1, mark.time=F)
```



Kaplan-Meier Analysis Example

To compare curves for different groups of subjects: Kaplan Meier Analysis according to one factor

To produce the Kaplan-Meier curves according to one factor. To compare survival curves according to a factor, log-rank test could be used. For the log-rank test "rho=0".

#Kaplan-Meier survival curves according to diabetes

```
> KM.diab <- survfit(Surv(data$fu,data$event)~ 1+data$diab)
> KM.diab
Call: survfit(formula = Surv(data$fu, data$event) ~ 1 + data$diab)

              n events median 0.95LCL 0.95UCL
data$diab=0 206     38     NA      NA      NA
data$diab=1  68     24     37      36      NA
```

Kaplan-Meier Analysis Example

HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

```
> KM.diab <- survfit(Surv(data$fu,data$event)~ 1+data$diab)
> KM.diab
Call: survfit(formula = Surv(data$fu, data$event) ~ 1 + data$diab)

                 n events median 0.95LCL 0.95UCL
data$diab=0 206     38     NA      NA      NA
data$diab=1  68     24     37      36      NA
```

```
                data$diab=0
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    1    206       2    0.990 0.00683        0.977        1.000
    2    202       4    0.971 0.01179        0.948        0.994
    6    178       2    0.960 0.01396        0.933        0.988
    8    160       2    0.948 0.01616        0.917        0.980
    9    148       2    0.935 0.01830        0.900        0.972
   10    146       2    0.922 0.02017        0.883        0.963
   12    132       2    0.908 0.02215        0.866        0.953
   13    130       6    0.866 0.02694        0.815        0.921
   17    110       2    0.851 0.02866        0.796        0.909
   23     88       2    0.831 0.03110        0.772        0.894
   27     72       2    0.808 0.03425        0.744        0.878
   35     48       4    0.741 0.04500        0.658        0.834
   37     42       2    0.705 0.04929        0.615        0.809
   38     40       2    0.670 0.05276        0.574        0.782
   42     36       2    0.633 0.05601        0.532        0.753

                data$diab=1
 time n.risk n.event survival std.err lower 95% CI upper 95% CI
    2     64       2    0.969  0.0217       0.9270        1.000
    5     56       2    0.934  0.0319       0.8737        0.999
    7     52       2    0.898  0.0395       0.8240        0.979
    9     50       2    0.862  0.0454       0.7778        0.956
   14     44       2    0.823  0.0511       0.7288        0.930
   18     38       2    0.780  0.0568       0.6760        0.900
   22     34       2    0.734  0.0621       0.6218        0.866
   24     24       2    0.673  0.0704       0.5481        0.826
   36     14       2    0.577  0.0872       0.4288        0.775
   37     12       2    0.481  0.0955       0.3255        0.709
   51      6       2    0.320  0.1123       0.1612        0.637
   60      4       2    0.160  0.0978       0.0484        0.530
```

Kaplan-Meier Analysis Example

HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

```
> KM.diab <- survfit(Surv(data$fu,data$event)~ 1+data$diab)
> KM.diab
Call: survfit(formula = Surv(data$fu, data$event) ~ 1 + data$diab)

                 n events median 0.95LCL 0.95UCL
data$diab=0 206     38     NA      NA      NA
data$diab=1  68     24     37      36      NA
```

Smallest time at which the survival probability drops to 0.5 (50%) or below.

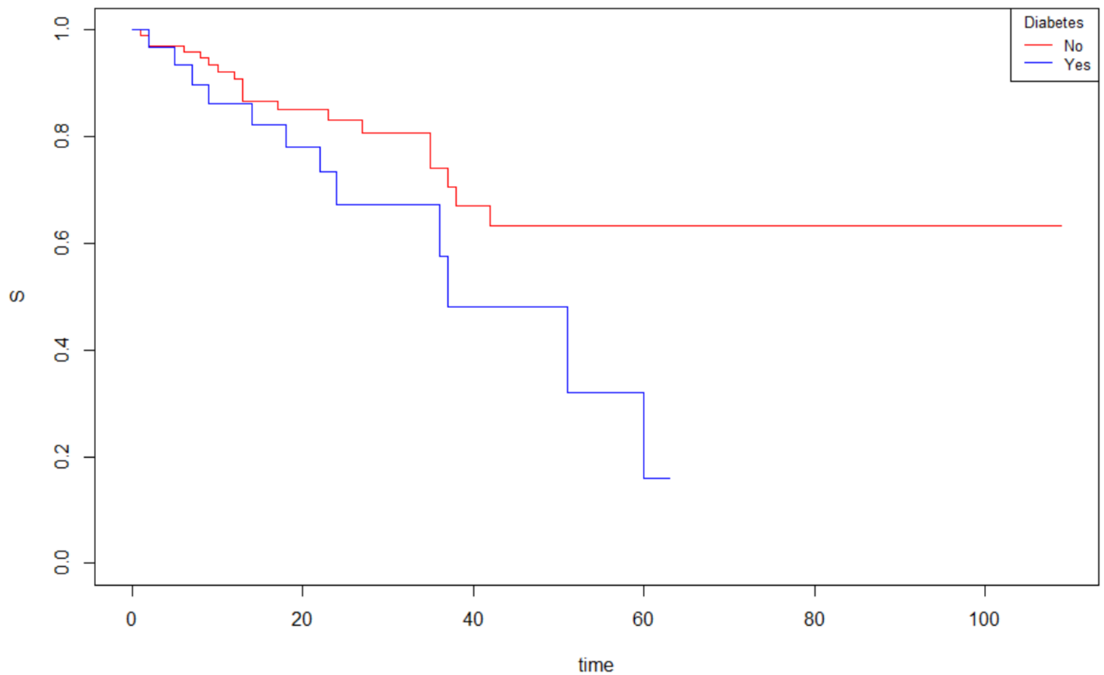
Median survival time = 37.

```
                data$diab=0
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   1    206       2    0.990 0.00683        0.977        1.000
   2    202       4    0.971 0.01179        0.948        0.994
   6    178       2    0.960 0.01396        0.933        0.988
   8    160       2    0.948 0.01616        0.917        0.980
   9    148       2    0.935 0.01830        0.900        0.972
  10    146       2    0.922 0.02017        0.883        0.963
  12    132       2    0.908 0.02215        0.866        0.953
  13    130       6    0.866 0.02694        0.815        0.921
  17    110       2    0.851 0.02866        0.796        0.909
  23     88       2    0.831 0.03110        0.772        0.894
  27     72       2    0.808 0.03425        0.744        0.878
  35     48       4    0.741 0.04500        0.658        0.834
  37     42       2    0.705 0.04929        0.615        0.809
  38     40       2    0.670 0.05276        0.574        0.782
  42     36       2    0.633 0.05601        0.532        0.753

                data$diab=1
time n.risk n.event survival std.err lower 95% CI upper 95% CI
   2     64       2    0.969  0.0217       0.9270        1.000
   5     56       2    0.934  0.0319       0.8737        0.999
   7     52       2    0.898  0.0395       0.8240        0.979
   9     50       2    0.862  0.0454       0.7778        0.956
  14     44       2    0.823  0.0511       0.7288        0.930
  18     38       2    0.780  0.0568       0.6760        0.900
  22     34       2    0.734  0.0621       0.6218        0.866
  24     24       2    0.673  0.0704       0.5481        0.826
  36     14       2    0.577  0.0872       0.4288        0.775
  37     12       2    0.481  0.0955       0.3255        0.709
  51      6       2    0.320  0.1123       0.1612        0.637
  60      4       2    0.160  0.0978       0.0484        0.530
```

Kaplan-Meier Analysis Example

#Plot survival curve according to diabetes

```
> plot(KM.diab,col=c("red","blue"),mark.time=F,ylim=c(0,1),xlab="time",ylab="S")
> legend("topright", title="Diabetes", legend=c("No", "Yes"), col=c("red", "blue"),
+        lty=1:1, cex=0.8)
```



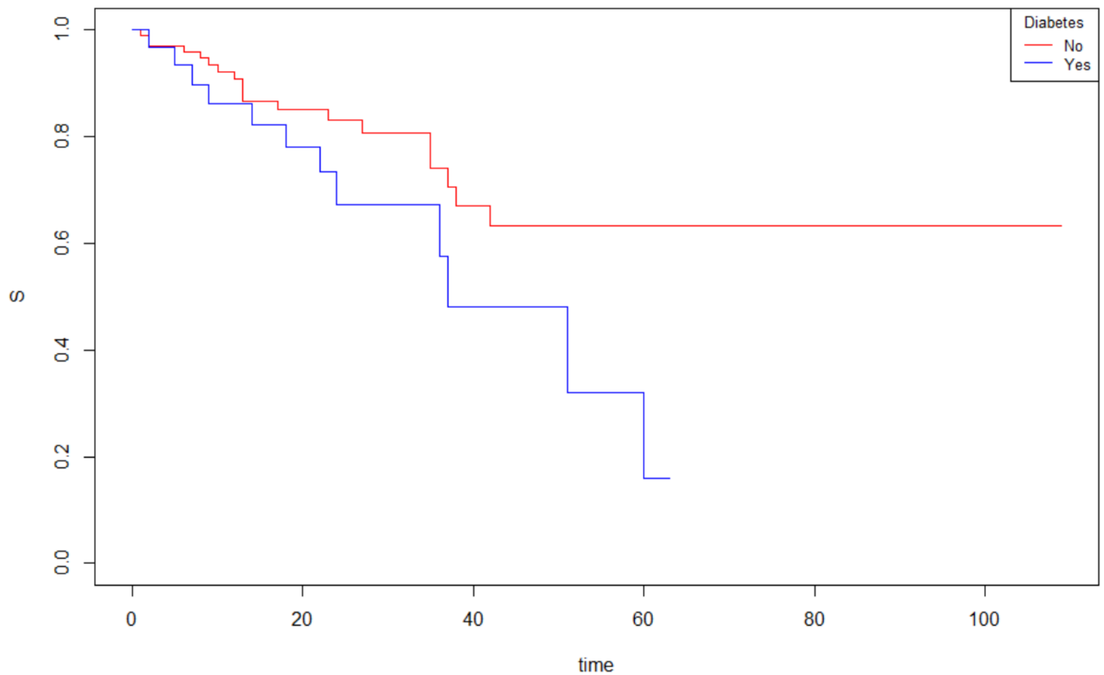
Kaplan-Meier Analysis Example

HEADS
PhD PROGRAMME IN HEALTH DATA SCIENCE

#Plot survival curve according to diabetes

```
> plot(KM.diab,col=c("red","blue"),mark.time=F,ylim=c(0,1),xlab="time",ylab="S")
> legend("topright", title="Diabetes", legend=c("No", "Yes"), col=c("red", "blue"),
+         lty=1:1, cex=0.8)
```

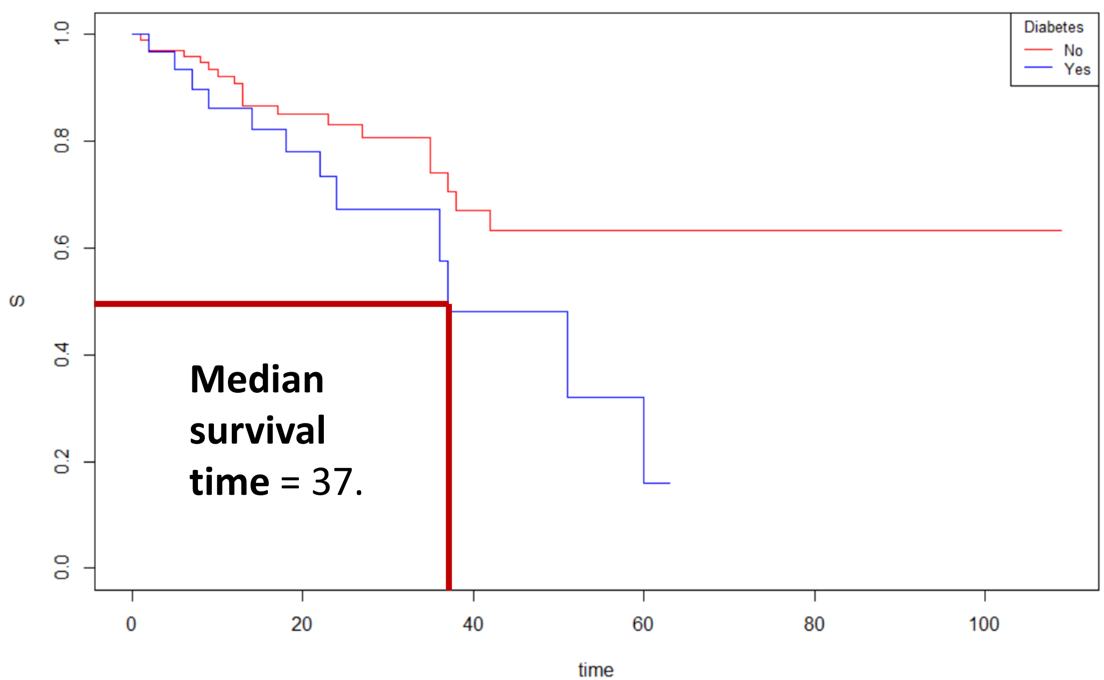
The Kaplan-Meier plot is now split into each of the levels of the categorical variable (2 groups in this case).



Kaplan-Meier Analysis Example

#Plot survival curve according to diabetes

```
> plot(KM.diab,col=c("red","blue"),mark.time=F,ylim=c(0,1),xlab="time",ylab="S")
> legend("topright", title="Diabetes", legend=c("No", "Yes"), col=c("red", "blue"),
+        lty=1:1, cex=0.8)
```



Kaplan-Meier Analysis Example

#Log-rank test

- Allows for comparison between groups.

- Possible to compute hand (based on Chi-square).

H_0: No difference between the groups.
H_1: The groups are different.

Kaplan-Meier Analysis Example

#Log-rank test

```
> plot(KM.diab,col=c("red","blue"),mark.time=F,ylim=c(0,1),xlab="time",ylab="S")
> legend("topright", title="Diabetes", legend=c("No", "Yes"), col=c("red", "blue"),
+        lty=1:1, cex=0.8)
> survdiff(Surv(data$fu,data$event)~ 1+data$diab,rho=0)
Call:
survdiff(formula = Surv(data$fu, data$event) ~ 1 + data$diab,
    rho = 0)

              N Observed Expected (O-E)^2/E (O-E)^2/V
data$diab=0 206       38     47.5      1.89      8.27
data$diab=1  68       24     14.5      6.19      8.27

 Chisq= 8.3  on 1 degrees of freedom, p= 0.004
```

Kaplan-Meier Analysis Example

#Log-rank test

```
> plot(KM.diab,col=c("red","blue"),mark.time=F,ylim=c(0,1),xlab="time",ylab="S")
> legend("topright", title="Diabetes", legend=c("No", "Yes"), col=c("red", "blue"),
+        lty=1:1, cex=0.8)
> survdiff(Surv(data$fu,data$event)~ 1+data$diab,rho=0)
Call:
survdiff(formula = Surv(data$fu, data$event) ~ 1 + data$diab,
    rho = 0)

              N Observed Expected (O-E)^2/E (O-E)^2/V
data$diab=0 206       38     47.5      1.89      8.27
data$diab=1  68       24     14.5      6.19      8.27

 Chisq= 8.3  on 1 degrees of freedom, p= 0.004
```

The log rank test here shows significant difference between the groups ($p < 0.05$).

Bibliografia

- David W. Hosmer DW, Lemeshow S, May S. (2008). **Applied Survival Analysis: Regression Modeling of Time-to-Event Data 2nd Edition**. New Jersey: John Wiley & Sons.
- Kleinbaum, DG and Klein M. (2012). **Survival Analysis – A Self- Learning Text, 3rd Edition**. New York: Springer-Verlag.
- Moore DFF. (2016). **Applied Survival Analysis Using R, 1st Edition**. New York: Springer-Verlag.
- Klein, J. and Moeschberger, M. (2003). **Survival Analysis - Techniques for Censored and Truncated Data**. New York: Springer-Verlag.
- Kalbfleisch, J. and Prentice, R. (2002). **The Statistical Analysis of Failure Time Data, 2nd Edition**. New York: Wiley.
- Cox, D. and Oakes, D. (1984). **Analysis of Survival Data**. London: Chapman & Hall.