

Survival analysis

Import the data

```
library(tidyverse)

## -- Attaching packages ----- tidyverse 1.2.1 --
## v ggplot2 3.0.0      v purrr  0.2.5
## v tibble  1.4.2      v dplyr  0.7.6
## v tidyr   0.8.1      v stringr 1.3.1
## v readr   1.1.1      v forcats 0.3.0

## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()

library(survival)
library(survminer)

## Loading required package: ggpubr
## Loading required package: magrittr
##
## Attaching package: 'magrittr'
##
## The following object is masked from 'package:purrr':
##
##   set_names
##
## The following object is masked from 'package:tidyr':
##
##   extract

library(readxl)
library(coin)

cancer <- read_excel("cancer_survival.xlsx")
cancer$event <- as.logical(cancer$event)
cancer$metastasised <- as.factor(cancer$metastasised)
head(cancer,3)

## # A tibble: 3 x 3
##   Following_Time event metastasised
##         <dbl> <lgl> <fct>
## 1          23 TRUE  no
## 2          47 TRUE  no
## 3          70 FALSE no

tail(cancer,3)

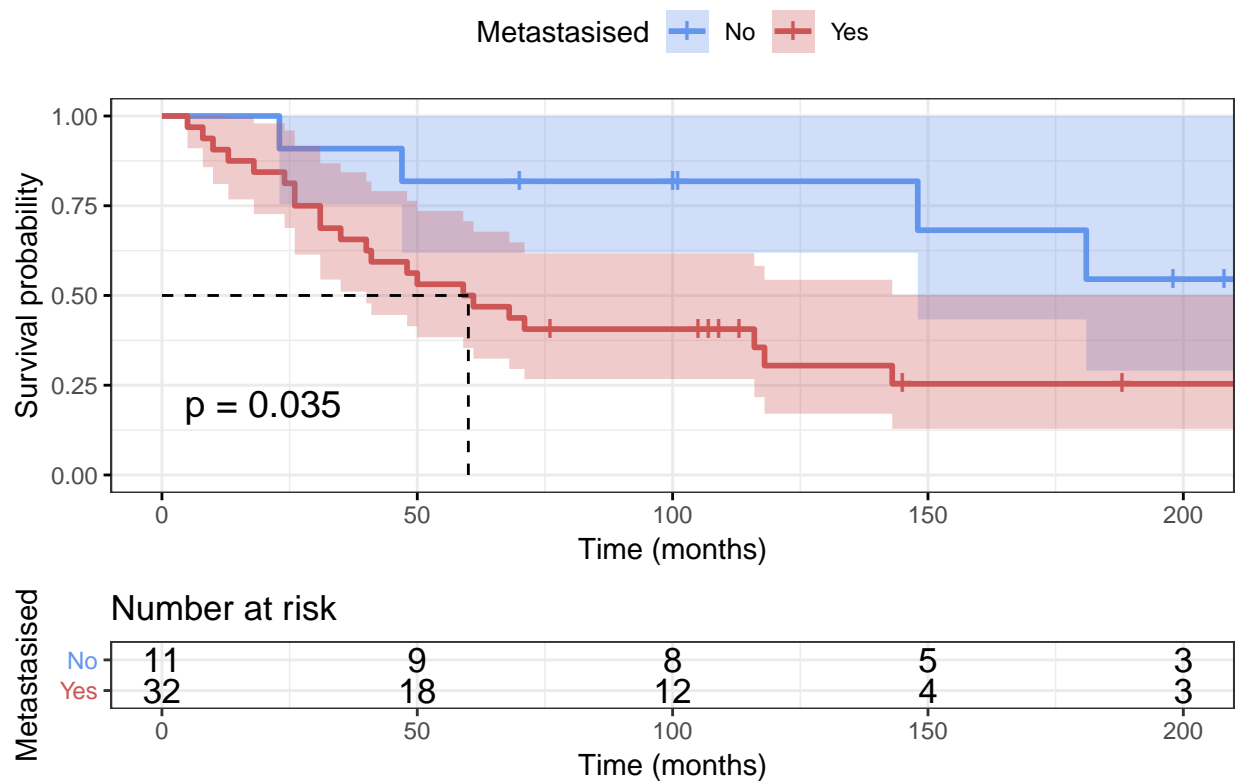
## # A tibble: 3 x 3
##   Following_Time event metastasised
##         <dbl> <lgl> <fct>
## 1          212 FALSE yes
## 2          217 FALSE yes
```

```
## 3          225 FALSE yes
```

Plot Kaplan-Meier function

```
sfit <- survfit(Surv(Following_Time,event)~metastasised,data=cancer)
ggsurvplot(sfit,
  conf.int=TRUE,
  pval=TRUE,
  legend.labs=c("No", "Yes"),
  legend.title="Metastasised",
  palette=c("cornflowerblue", "indianred3"),
  title="Kaplan-Meier Curve for breast cancer survival",
  xlab = "Time (months)",
  risk.table = T,
  risk.table.height = 0.25, # Useful to change when you have multiple groups
  surv.median.line = "hv", # add the median survival pointer.
  ggtheme = theme_bw())
```

Kaplan-Meier Curve for breast cancer survival



```
##logrank Test
```

```
logrank_test(Surv(Following_Time, event) ~ metastasised, data = cancer, distribution = "exact")
```

```
##
```

```
## Exact Two-Sample Logrank Test
```

```
##
```

```
## data: Surv(Following_Time, event) by metastasised (no, yes)
```

```
## Z = 2.2773, p-value = 0.02123
## alternative hypothesis: true theta is not equal to 1
```