

Angewandte Datenanalyse mit R

Tag 4 - Überlebenszeitanalysen

Andreas Mock

Abteilung für Medizinische Onkologie & Abteilung für Translationale Medizinische Onkologie, Nationales
Centrum für Tumorerkrankungen (NCT) Heidelberg

Sommersemester 2019



HEIDELBERG
UNIVERSITY
HOSPITAL



GERMAN
CANCER RESEARCH CENTER
IN THE HELMHOLTZ ASSOCIATION



DKTK

German Cancer
Consortium

Überlebenszeitanalyse

Kaum eine klinische Studie kommt ohne Überlebenszeitanalysen aus. Deshalb sollte die sachkundige Durchführung und Interpretation dieser zum Rüstzeug jedes Mediziners gehören.

Terminologie

- ▶ Hazard Ratio (HR)
- ▶ Censor
- ▶ univariat
- ▶ multivariat

Überlebenszeitanalyse

Terminologie

- ▶ **Hazard Ratio (HR):** Relative Wahrscheinlichkeit zwischen den Gruppen innerhalb eines Zeitintervalls (z.B. innerhalb eines Monats oder Jahres) ein Event zu haben (z.B. Tod oder Progress)
- ▶ **Censor:** 1 = Event eingetroffen (z.B. Tod oder Progress), 0 = Event noch nicht eingetroffen
- ▶ **univariat:** Es wird nur der Einfluss eines Faktors (z.B. Geschlecht) auf den Endpunkt (z.B. Gesamtüberleben) untersucht.
- ▶ **multivariat:** Es wird nur der Einfluss mehrerer Faktoren auf den Endpunkt untersucht.

Überlebenszeitanalyse

Pakete

- ▶ `survival`: Paket für statistischen Berechnungen
- ▶ `survminer`: Paket für Visualisierung der Kaplan-Meier-Kurven auf Basis von `ggplot2`
- ▶ `broom`: Paket für Ergebnistabellen

Datensatz

Wir werden wieder mit dem `hnscc`-Datensatz arbeiten.

Überlebenszeitanalyse

Für eine vergleichende Überlebenszeitanalysen benötigen wir drei Informationen:

1. die Zeit bis zu einem Event (z.B. Tod oder Progress)
2. der Censor (binäre Info ob Event eingetroffen ist, z.B. 1 = tod, 0 = lebend)
3. den zu untersuchenden Einflussfaktor (diskrete oder kontinuierliche Daten)

```
# 1. Zeit bis zum Tod  
summary(hnsccl$days_to_death)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.     NA's  
##      0.0   218.8   443.0   789.0   999.2  6416.0         1
```

```
# 2. Censor  
table(hnsccl$vital_status)
```

```
##  
## DECEASED    LIVING  
##      116      163
```

```
# 3. den zu untersuchenden Einflussfaktor  
colnames(hnsccl)[-c(1,4,11)]
```

```
## [1] "age"           "alcohol"        "gender"         "neoplasm_site"  
## [5] "grade"        "pack_years"     "tabacco_group"  "tumor_stage"
```

Datentransformation für Überlebenszeitanalyse

1. Der Censor muss ein **numerischer Vektor** aus 0 (=Event nicht eingetreten) und 1 (=Event eingetreten) sein.
2. Die Zeit bis zu einem Event (z.B. Tod) muss **numerisch** sein.

Diese Datentransformation entspricht der Übung 1 des heutigen Tages. Hierbei werden wir für die weitere Analyse ein eigenes Objekt `hnscc_survival` erstellen.

Erstellung eines survival Objektes

Für Überlebenszeitanalysen in R gilt es von so genannten `survival` Objekten gebrauch zu machen. Diese sind Teil des `survival` Pakets.

Die Syntax ist straightforward:

```
head(Surv(hnscs_survival$OS, hnscs_survival$Censor))
```

```
## [1] 461 415 1134 276 248+ 190+
```

Kaplan-Meier-Kurve

Der Klassiker der Visualisierung von Überlebenszeitanalysen ist die so genannte Kaplan-Meier-Kurve.

Wichtig: Überlebenszeitanalysen können auch für kontinuierliche Einflussfaktoren durchgeführt werden. Hier macht natürlich eine Visualisierung mittels Kaplan-Meier-Kurve keinen Sinn.

Für diese müssen wir noch die Gruppenzugehörigkeit angeben, und das Objekt für den Plot mit `survfit` herstellen:

```
median(hnscs_survival$age)
```

```
## [1] 61
```

```
fit <- survfit(Surv(OS, Censor)~age>61, data=hnscs_survival)
fit
```

```
## Call: survfit(formula = Surv(OS, Censor) ~ age > 61, data = hnscs_survival)
```

```
##
```

```
##      1 observation deleted due to missingness
```

```
##           n events median 0.95LCL 0.95UCL
```

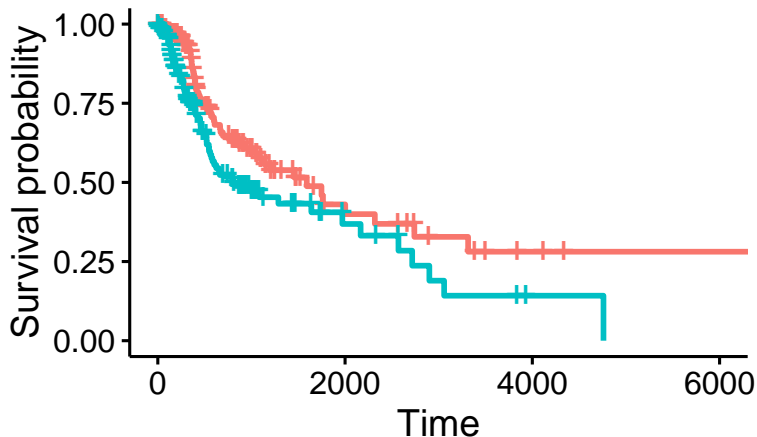
```
## age > 61=FALSE 139      50   1591    1037    3314
```

```
## age > 61=TRUE  139      66    822     572    2166
```


Kaplan-Meier-Kurve

```
ggsurvplot(fit, hnscc_survival)
```

Strata + age > 61=FALSE + age > 61=TRUE



Univariate Überlebenszeitanalyse

```
library(broom)
```

```
# Beispiel 1: Alter dichotomisiert nach > Median (= 61 Jahre)
```

```
tidy(coxph(Surv(OS, Censor)~age>61, data=hnscc_survival),  
      exponentiate = TRUE)
```

```
## # A tibble: 1 x 7
```

```
##   term          estimate std.error statistic p.value conf.low conf.high  
##   <chr>          <dbl>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>  
## 1 age > 61TRUE      1.54     0.189      2.27  0.0229     1.06     2.23
```

```
# Beispiel 2: Alter als kontinuierliche Variable
```

```
tidy(coxph(Surv(OS, Censor)~age, data=hnscc_survival),  
      exponentiate = TRUE)
```

```
## # A tibble: 1 x 7
```

```
##   term estimate std.error statistic p.value conf.low conf.high  
##   <chr>   <dbl>    <dbl>    <dbl>   <dbl>   <dbl>   <dbl>  
## 1 age    1.02   0.00827      2.69 0.00718     1.01     1.04
```

Estimate = HR

Univariate Überlebenszeitanalyse

```
# Beispiel 3: Tobacco Group
```

```
table(hnscs_survival$tabacco_group)
```

```
##  
## Current reformed smoker for < or = 15 years  
##                               81  
##      Current reformed smoker for > 15 years  
##                               49  
##                               Current smoker  
##                               90  
##                               Lifelong Non-smoker  
##                               52
```

```
tidy(coxph(Surv(OS, Censor)~tabacco_group, data=hnscs_survival),  
      exponentiate = TRUE)
```

```
## # A tibble: 3 x 7  
##   term                estimate std.error statistic p.value conf.low conf.high  
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>  
## 1 tabacco_groupCur~  0.572    0.290    -1.93    0.0542    0.324    1.01  
## 2 tabacco_groupCur~  1.14     0.234     0.551    0.582     0.719    1.80  
## 3 tabacco_groupLif~  0.482    0.307    -2.38    0.0173    0.264    0.879
```

Multivariate Überlebenszeitanalyse

In der multivariaten Überlebenszeitanalyse kombinieren wir die Einflussfaktoren zusammen, die univariat signifikant waren. Wir wollen herausfinden, ob diese **unabhängige** Einflussfaktoren sind.

```
# Beispielmmodel: Alter und Tobacco group
tidy(coxph(Surv(OS, Censor)~age + tobacco_group, data=hnscc_survival),
      exponentiate = TRUE)
```

```
## # A tibble: 4 x 7
```

##	term	estimate	std.error	statistic	p.value	conf.low	conf.high
##	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
## 1	age	1.02	0.00931	2.58	0.00986	1.01	1.04
## 2	tobacco_groupCur~	0.520	0.292	-2.23	0.0255	0.293	0.923
## 3	tobacco_groupCur~	1.22	0.235	0.838	0.402	0.768	1.93
## 4	tobacco_groupLif~	0.553	0.309	-1.92	0.0552	0.301	1.01