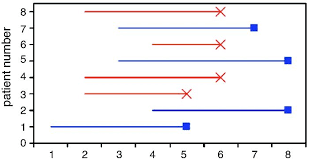
Dealing with dates in R (survival time calculation)

Hyemi Choi

2020 9 7

## R을 이용한 생존시간 계산

시간 관측 값이 *시점*과 *종점* 으로 주어질 때 생존시간 계산 

* base R
* lubridate package

### lubridate Package

* R에서 날짜 자료 다루는 패키지
* Garrett Grolemund 와 Hadley Wickham 외 8명이 만듬
* tidyverse에 포함되어 있음

library(tidyverse)  
library(lubridate)

* ymd, mdy, dmy 함수 : dates 객체 생성
* year, month, day 함수 : dates 객체로부터 부분 정보 추출

ymd("2020/09/07")

## [1] "2020-09-07"

mdy("09/07/2020")

## [1] "2020-09-07"

dmy("07/09/2020")

## [1] "2020-09-07"

d1 <- ymd("2020/09/07")  
year(d1)

## [1] 2020

### 관측 시작 시점과 마지막 시점이 주어진 생존자료

1. R의 date 형식인지 확인하고 아니면 date 형식으로 변환
2. 생존시간 계산

**R을 이용한 생존시간 계산**

## Example

* sx\_date : 수술 날짜 (surgery date )
* last\_fup\_date : 마지막 추적 관측 날짜 (last follow-up date)

date\_ex <-   
 tibble(  
 sx\_date = c("2007-06-22", "2004-02-13", "2010-10-27"),   
 last\_fup\_date = c("2017-04-15", "2018-07-04", "2016-10-31")  
 )  
  
glimpse(date\_ex)

## Rows: 3  
## Columns: 2  
## $ sx\_date <chr> "2007-06-22", "2004-02-13", "2010-10-27"  
## $ last\_fup\_date <chr> "2017-04-15", "2018-07-04", "2016-10-31"

*NOTE* - 두 변수는 문자변수 date 형식으로 변환 필요 - 두 날짜의 차이 계산한 후 숫자 형식으로 변환

### base R 이용

* as.Date

#### Formatting dates

(date\_r<- date\_ex %>%   
 mutate(  
 sx\_date = as.Date(sx\_date, format = "%Y-%m-%d"),   
 last\_fup\_date = as.Date(last\_fup\_date, format = "%Y-%m-%d")   
 ))

## # A tibble: 3 x 2  
## sx\_date last\_fup\_date  
## <date> <date>   
## 1 2007-06-22 2017-04-15   
## 2 2004-02-13 2018-07-04   
## 3 2010-10-27 2016-10-31

* 주어진 날짜 자료가 m/d/Y 형식이면 format = "%m/%d/%Y"로 입력
* date format symbols 참조 <https://www.statmethods.net/input/dates.html>

#### Calculating survival times

* difftime : 경과 시간 일 단위 (units=“days”)
* as.numeric: 숫자 형식으로 변환
* /365.25: 년 단위로 변환

date\_r %>%   
 mutate(  
 os\_yrs =   
 as.numeric(  
 difftime(last\_fup\_date,   
 sx\_date,   
 units = "days")) / 365.25  
 )

## # A tibble: 3 x 3  
## sx\_date last\_fup\_date os\_yrs  
## <date> <date> <dbl>  
## 1 2007-06-22 2017-04-15 9.82  
## 2 2004-02-13 2018-07-04 14.4   
## 3 2010-10-27 2016-10-31 6.01

### lubridate package 이용

* ymd

#### Formatting dates

(date\_l <- date\_ex %>%   
 mutate(  
 sx\_date = ymd(sx\_date),   
 last\_fup\_date = ymd(last\_fup\_date)  
 ))

## # A tibble: 3 x 2  
## sx\_date last\_fup\_date  
## <date> <date>   
## 1 2007-06-22 2017-04-15   
## 2 2004-02-13 2018-07-04   
## 3 2010-10-27 2016-10-31

* base R 에서와 달리 분리자가 필요없음

#### Calculating survival times

* %--% 연산자 : 시간 구간 차이
* as.duration : 초 단위 경과시간
* dyears(1): 년 단위로 변환

date\_l %>%   
 mutate(  
 os\_yrs =   
 as.duration(sx\_date %--% last\_fup\_date) / dyears(1)  
 )

## # A tibble: 3 x 3  
## sx\_date last\_fup\_date os\_yrs  
## <date> <date> <dbl>  
## 1 2007-06-22 2017-04-15 9.82  
## 2 2004-02-13 2018-07-04 14.4   
## 3 2010-10-27 2016-10-31 6.01

### 과제

각자 태어나서 지금까지 몇 일 살았는지 base R과 lubridate 패키지를 이용하여 계산하기