Tölfræði fyrir almenning

Sigurbjörg Anna Guðnadóttir

2020-08-17

Contents

Þeir pakkar sem við notum eru:

```
library(tidyverse)
library(survival)
library(flexsurv)
library(survminer)
library(arsenal)
library(table1)
```

Chapter 1

Línuleg aðhvarfsgreining

Í línulegri aðhvarfsgreiningu gildir að útkoman þarf að vera línuleg breyta. Það gildir ekki um skýribreyturnar.

Þeir pakkar sem við notum eru:

```
library(tidyverse)
library(epitools)
library(arsenal)
library(ggpubr)
library(table1)
```

1.1 Lýsandi tölfræði

Western Collaborative Group Study (WCGS) gagnasafnið er í epitools-pakkanum. Gagnasafnið byggir á rannsókn sem var með meginmarkmið að skoða tengsl persónuleikagerðar og hjartasjúkdóma. Viðfangsefnin voru 3154 karlmenn á aldrinum 39 - 59 ára, þeim var fylgt eftir í allt að 10 ár þangað til þeir fengu hjartasjúkdóm, létust, þeir urðu 70 ára eða eftirfylgni lauk af öðrum ástæðum. Viðfangsefnin komu inn í rannsóknina á árunum 1960-1961. Í þessum kafla og næstu ætlum við að skoða tengsl kólesteróls og reykinga.

Chapter 2

Lifunargreining

Í þessum kafla munun við styðjast við Western Collaborative Group Study (WCGS) gagnasafnið sem er í epitools- pakkanum. Þetta eru gögn úr rannsókn þar sem meginmarkmið hennar var að skoða tengsl persónuleikagerðar og hjartasjúkdóma og því eru hjartasjúkdómar aðalútkomubreytan okkar. Viðfangsefnin voru 3154 karlmenn á aldrinum 39 - 59 ára, þeim var fylgt eftir í allt að 10 ár þangað til þeir fengu hjartasjúkdóm, eða létust, eða þeir urðu 70 ára eða eftirfylgni lauk af öðrum ástæðum. Viðfangsefnin komu inn í rannsóknina á árunum 1960-1961.

Við ætlum að skoða áhrif mismunandi persónuleikagerða A og B á hjartaáföll með og án skýribreytna. A og B flokkunin vísar í hvernig fólk meðhöndlar streitu og álag. Þeir sem eru í A hópi eru með meira keppnisskap, óþolinmóðari og árásagjarnari en þeir sem eru í hópi B. Þær skýribreytur sem við munum skoða eru aldur, magn kólesteról í blóði, blóðþrýstings, reykinga og BMI stuðull.

2.1 Skoðun lifunargagna

2.1.1 Sækjum gögnin og lögum til

```
data(wcgs,package = "epitools")
wcgs <- as_tibble(wcgs)

wcgs <- wcgs %>%
  mutate(surv_time_y = time169 / 365.24,
         agec = age0 - 46,
         cholmmol = chol0 / 39,
         sbp10 = sbp0 / 10,
         dibpat = factor(dibpat0, levels = 0:1, labels = c("B", "A")),
         smoker = factor(1 * (ncigs0 > 0), levels = c(0, 1), labels = c("No", "Yes")),
```

```
        bmi = (weight0 * 0.454) / ((height0 * 2.54)/100)^2,
        bmiq3 = cut(bmi, breaks = quantile(bmi, seq(0, 1, 1/3)),
                    include.lowest = T, right = F)
        )
wcgs_dat <- wcgs %>%
  select(id, surv_time = time169, surv_time_y, status=chd69,  agec, cholmmol, sbp10, sm
  filter(complete.cases(.)) %>% mutate(statusf=factor(status,levels=0:1,labels = c("No

wcgs_dat$dibpat <- relevel(wcgs_dat$dibpat, "A")
```

Tíminn time169 er í dögum en við reiknum nýja breytu sem mælir tímann í árum.
Við búum til nýja aldursbreytu þar sem við erum búin að staðla hana m.v. meðaldurinn í hópnum. Meðalaldurinn er 46.28 og drögum við því 46 frá aldrinum. Það mun auðvelda túlkun á líkönum.
!!! ATH af hverju deilum við með 10 og 39?.
Breytan dibpat0 eru persónuleikagerðirnar, við setjum hana sem flokkabreytu (e. factor).
Við skilgreinum reykingamann þann sem reykir amk 1 sígaréttu á dag og höfum breytuna sem flokkabreytu.
Til þess að reikna BMI þá þurfum við að breyta hæðinni í metra og þyngdinni í kíló. Útbúum einnig flokkabretytu fyrir bmi þar sem við skiptum henni í 3 jafna hluta. Breytan chd69 segir til um hvort karlmennirnir fengu hjartasjúkdóm eða ekki og setjum við hana því sem "status", það hvort atburður hafi átt sér stað eða ekki er oft kallað "status".

Við notum bara þá einstaklinga sem hafa allar breyturnar sem við ætlum að skoða, aðra fjarlægjum við úr gagnasafninu.

2.1.2 Tökum létta skoðun á gögnunum

Fyrsta skrefið er alltaf að skoða gögnin og sjá hvað einfaldur reikningur gefur okkur. Við skoðum gögnin miða við ár sem tímalengd.

Helstu tölur

```
# Hversu margir eru í safninu og í hvorum hópi fyrir sig?
rownames <- c("Allir", "A", "B")
fj_t <- dim(wcgs_dat)[1]
fj_g <- wcgs_dat %>%
  group_by(dibpat) %>%
  count()

fj <- rbind(fj_t,fj_g[[1,2]],fj_g[[2,2]])

# Hversu margir fengu hjartaáfall í heildina og í hvorum hópi fyrir sig?
```

```
st_t <- sum(wcgs_dat$status)
st_g <- wcgs_dat %>%
  group_by(dibpat) %>%
  summarise(tidni = sum(status))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
st <- rbind(st_t,st_g[[1,2]],st_g[[2,2]])

# Hver er eftirfylgnitíminn og lambda fyrir all hópana?

ef_t <- wcgs_dat %>%
  summarise(sum_time = sum(surv_time_y),lambda = sum(status) / sum(surv_time_y))

ef_g <- wcgs_lambda <- wcgs_dat %>%
  group_by(dibpat) %>%
  summarise(sum_time = sum(surv_time_y),lambda = sum(status) / sum(surv_time_y))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
ef <- rbind(round(ef_t[[1]],2), round(ef_g[[1,2]],2), round(ef_g[[2,2]],2))
lambda <- rbind(round(ef_t[[2]],4), round(ef_g[[1,3]],4), round(ef_g[[2,3]],4))

# Setjum í eina töflu
stats <- as_tibble(cbind(rownames,fj,st,ef,lambda))

# !!! ATH betra útlit
stats <-stats %>% rename( Hópar= rownames, Heildarfjöldi = V2,  Tilfelli = V3, Eftirfylgnitími =
stats
```

```
## # A tibble: 3 x 5
##   Hópar Heildarfjöldi Tilfelli Eftirfylgnitími lambda
##   <chr> <chr>         <chr>    <chr>           <chr>
## 1 Allir 3140          255      23072.54        0.0111
## 2 A     1583          177      11330.47        0.0156
## 3 B     1557          78       11742.07        0.0066
```

Eftirfylgnitíminn er sá tími sem einstaklingur er í rannsókninni, hámark 10 ár í þessari rannsókn. Lambda er fjöldi atburða á tímaeiningu, í þessu tilfelli er það eitt ár. Svo 0.0111 eða 1.11 % er eins árs meðaláhætta fyrir heildarhópinn.

Áhættuhlutfallið Áhættuhlutfallið er hlutfallið af líkunum á því að atburðurinn gerist í meðferðarhópnum á móti líkunum á því að atburðurinn gerist í viðmiðunarhópnum. Við lítum á persónuleikagerð A sem meðferðarhópinn og persónuleikagerð B sem viðmiðunarhópinn.

```
hz <- wcgs_lambda$lambda[2]/wcgs_lambda$lambda[1]
```

Áhættuhlutfallið er 0.43 og því er einstaklingur með persónuleikagerð A í -57%

meiri áhættu til að fá hjartaáfall ef allir aðrir þættir eru eins.

Af þessum gildum höfum við mestan áhuga áhættunni (þ.e. lambda) fyrir hvora persónuleikagerð fyrir sig og áhættuhlutfallinu. Fallið *flexsurvreg* úr pakkanum *flexsurv* hentar vel til þess að reikna það. Einnig notum við fallið *Surv* úr pakkanum *survival*.
Fallið *Surv* er notað til að útbúa breytu af gerðinni lifunarhlutur (e. survival object). Það tekur inn í sig tvær breytur; annars vegar hversu langur tími leið fram að atburði eða skerðingu og hins vegar hvort atburður eða skerðing átti stað á þeim tímapunkti. Þeir sem hafa ekki fengið hjartaáfall eru með skerðingu.

Líkanið sem við köllum VL_0 er notað til að reikna lambda fyrir allan hópinn en VL_1 til þess að reikna lambda fyrir hópa A og B.

```
VL_0 <- flexsurvreg(Surv(surv_time_y, status) ~ 1,data=wcgs_dat,dist="exponential")
VL_1 <- flexsurvreg(Surv(surv_time_y, status) ~ dibpat,data=wcgs_dat,dist="exponential"

lambda_0 <- VL_0$res[1,1]
lambda_a <- VL_1$res[1,1]
lambda_b <- exp(VL_1$res[2,1]+log(VL_1$res[1,1]))


lambda_2 <- rbind(round(lambda_0,4), round(lambda_a,4), round(lambda_b,4))

# Setjum í eina töflu
stats2 <- as_tibble(cbind(rownames, lambda, lambda_2))

# !!! ATH betra útlit
stats2 <- stats2 %>% rename( Hópar= rownames, Lambda_handreiknað = V2,  Lambda_með_fal
stats2
```

```
## # A tibble: 3 x 3
##   Hópar Lambda_handreiknað Lambda_með_falli
##   <chr> <chr>              <chr>
## 1 Allir 0.0111             0.0111
## 2 A     0.0156             0.0156
## 3 B     0.0066             0.0066
```

Getum líka reiknað áhættuhlutfallið

```
hz_2 <- exp(VL_1$res[2,1])
```

Áhættuhlutfallið er 0.43 sem er það sama og við fengum með handreikningi.
«««< HEAD

Chapter 3

Veldisvísisfallið

3.1 Tafla 1

Þar sem við gerum yfirleitt líkön með skýribreytum þá er gott að útbúa töflu 1 til þess að fá tilfinningu fyrir dreifingu þeirra. Til eru ýmsar skipanir til að gera þessa töflu, meðal annars *tableyby* í pakkanum *arsenal*. Gott er að nefna breyturnar fyrst með skiljanlegum heitum. Notum til þess *label* úr pakkanum table1

```
label(wcgs_dat$dibpat) <- "Hegðunarhópur"
label(wcgs_dat$agec) <- "Aldur"
label(wcgs_dat$cholmmol) <- "Kólestról"
label(wcgs_dat$sbp10) <- "Blóðþrýstingur"
label(wcgs_dat$smoker) <- "Reykingar"
label(wcgs_dat$bmi) <- "BMI"
label(wcgs_dat$arcus0) <- "Hornhimnubogi" ## !!! ATH er þetta rétt íslenska?
```

```
tab1 <- tableby(dibpat ~ bmi + agec + cholmmol + sbp10 + smoker + bmi + arcus0, data= wcgs_dat)
summary(tab1, text = TRUE)
```

```
##
##
## |               |   A (N=1583)    |   B (N=1557)    | Total (N=3140)  | p value|
## |:--------------|:---------------:|:---------------:|:---------------:|-------:|
## |BMI            |                 |                 |                 |   0.130|
## |-  Mean (SD)   | 24.609 (2.600)  | 24.470 (2.531)  | 24.540 (2.567)  |        |
## |-  Range       | 11.202 - 37.285 | 15.676 - 38.986 | 11.202 - 38.986 |        |
## |Aldur          |                 |                 |                 | < 0.001|
## |-  Mean (SD)   |  0.769 (5.697)  | -0.228 (5.282)  |  0.275 (5.517)  |        |
## |-  Range       | -7.000 - 13.000 | -7.000 - 13.000 | -7.000 - 13.000 |        |
## |Kólestról      |                 |                 |                 |   0.001|
```