

Úloha 3

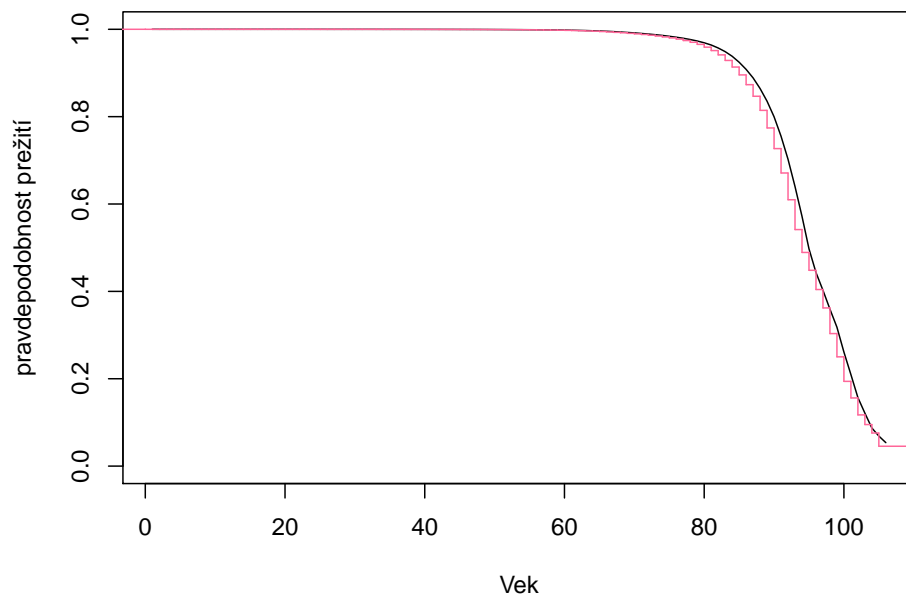
Analýza cenzorovaných dat

Michaela Minaříková

1 Kaplan Meier a aktuárský odhad

V datech máme proměnné i - věk, D_i - počet zemřelých ve věku i a Y_i - počet jedinců, kteří na počátku roku vstoupili do věkové skupiny i .

Nejdříve jsme pro odhad funkcí přežití použili Kaplan-Meierův odhad a aktuárský odhad pomocí funkce lifetab. Grafické výstupy jsou v obrázku 1.



Obrázek 1: Černá linka značí odhad pomocí funkce lifetab, růžová KM.

Problém použití obou metod spočívá v době pozorování - pozorujeme celou populaci pouze jeden rok, nikoliv celých 106 let. Proto kumulovaná proměnná počtu lidí nedává smysl, tím pádem jsou pravděpodobnosti přežití významně nadhodnoceny.

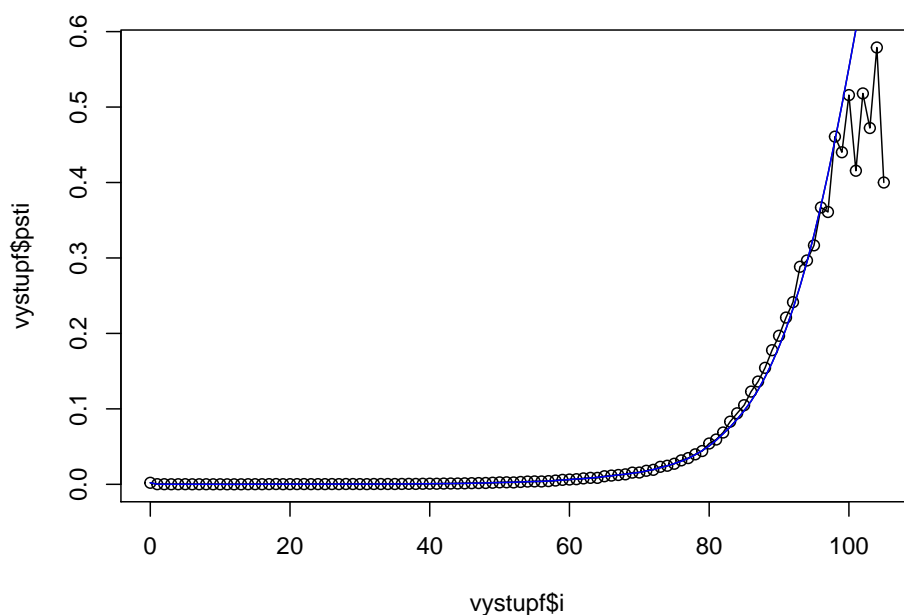
2 Správný odhad funkce přežití

Nechť X je n.v. délky života. Pro výpočet funkce přežití z úmrtnostních tabulek je nejdříve potřeba spočítat pravděpodobnost úmrtí ve věku x :

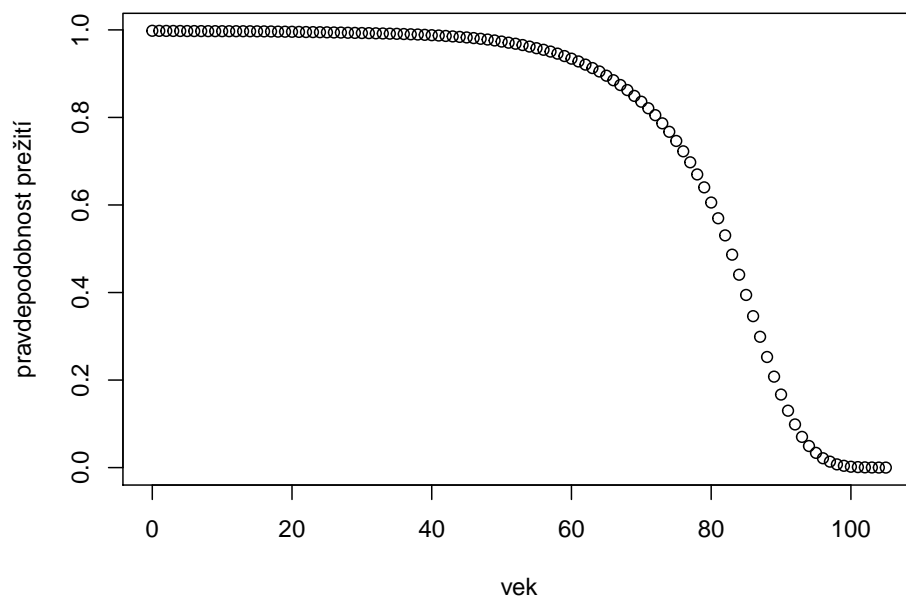
$$q_x = \frac{D_x}{Y_x}$$

$$\mathbb{P}(X = x) = q_x * p_{x-1} * \dots * p_0,$$

kde p_x značí pravděpodobnost přežití do věku $x + 1$ za podmínky, že se jedinec dožil věku x . Takto napočítané pravděpodobnosti úmrtí jsou myšlenkově totožné s výpočtem pravděpodobnosti úmrtí ČSÚ, rozdíl je ve věku 0, kde není splněný předpoklad rovnoměrnosti úmrtí během roku (nejvíce dětí roku 0 umírá krátce po porodu). Poté asi od věku 65 let už není tolik dat a proto ČSÚ využívá vyhlazování hrubě napočtených pravděpodobností. Proto jsem udělala srovnání jen graficky, v obrázku 2. V obrázku 3 je napočtená funkce přežití z daných pravděpodobností úmrtí.



Obrázek 2: Modrá linka značí psti úmrtí z ÚT, černá napočtené psti úmrtí z dat.



Obrázek 3: Pravděpodobnost přežití dle napočtených pstí úmrtí výše.

```
load("D:/Škola/Mgr/III. semestr/Analýza cenzorovaných dat/3.cviko/mort.RData")
library(KMSurv)

## ženy
vystupf <- mort.f

vystupf$y_i <- rev(cumsum(rev(vystupf$Yi)))
vystupf$pomer <- vystupf$Di/vystupf$y_i
vystupf$Li <- cumsum(vystupf$pomer)
vystupf$Si <- cumprod(1-vystupf$pomer)
vystupf$scenz <- vystupf$Yi-vystupf$Di
aktuarf <- lifetab(tis=c(vystupf$i,106),ninit = vystupf$y_i[1],
nlost = vystupf$scenz, nevent = vystupf$Di)

###muži
vystupm <- mort.m

vystupm$y_i <- rev(cumsum(rev(vystupm$Yi)))
vystupm$pomer <- vystupm$Di/vystupm$y_i
vystupm$Li <- cumsum(vystupm$pomer)
```

```

vystupm$Si <- cumprod(1-vystupm$pomer)
vystupm$scenz <- vystupm$Yi-vystupm$Di
aktuarm <- lifetab(tis=c(vystupm$i,106),ninit = vystupm$y_i[1],
nlost = vystupm$scenz, nevent = vystupm$Di)

##správný odhad
vystupf$psti <- vystupf$Di/vystupf$Yi
vystupm$psti <- vystupm$Di/vystupm$Yi

#POROVNÁNÍ
umrtnost_cr_z_2013 <- read_excel("C:/Users/misa1/Downloads/umrtnost_cr_z_2013.xlsx",
+ skip = 2)
dim(umrtnost_cr_z_2013)

lines(vystupf$i,vystupf$psti)
par(new=FALSE)
lines(umrtnost_cr_z_2013$'věk age',umrtnost_cr_z_2013$qx, col='blue')

plot(vystupf$i,cumprod(1-vystupf$psti), xlab = "věk",ylab = "pravděpodobnost přežití")

plot(aktuarf$surv,ylim = c(0,1), type="l", xlab = "Věk", ylab = "pravděpodobnost přežití")
plot(stepfun(vystupf$i, c(1,vystupf$Si)) ,col = "#FF6699",
do.points=FALSE, ylim = c(0,1), type="l",add = TRUE)

```