

Ćwiczenia 6

1. (S) Śmierć na raka po ekspozycji na dany czynnik rakotwórczy zmierzono w dwóch grupach szczurów. Grupa 1 otrzymała inną dawkę czynnika niż grupa 2. Rejestrujemy czas życia od ekspozycji. Jeśli szczur przeżył do końca eksperymentu, albo zmarł z innej przyczyny, to czas jest uważany za cenzurowany. Uzyskano następujące czasy przeżycia: Grupa 1: 143, 165, 188, 188, 190, 192, 206, 208, 212, 216, 220, 227, 230, 235, 246, 265, 303, 216⁺, 244⁺; Grupa 2: 142, 157, 163, 198, 205, 232, 232, 232, 233, 233, 233, 233, 239, 240, 261, 280, 280, 295, 295, 323, 204⁺, 344⁺. Znajdź estymator KM funkcji przeżycia. Oszacuj średni czas przeżycia oraz medianę tego czasu dla każdej grupy. Narysuj estymatory KM dla obu grup na jednym wykresie.
2. Zbiór danych `cancer` z pakietu `survival` zawiera dane na temat czasów przeżycia pacjentów z zaawansowanym rakiem płuc. Zmienna `status` zawiera informacje czy dane są cenzorowane. Znajdź estymator KM ze względu na płeć (narysuj go). Jaki jest średni czas przeżycia kobiet i mężczyzn, ile wynosi mediana tego czasu? Czy różnica pomiędzy krzywymi czasów przeżycia jest istotna? Stwórz odpowiedni model zależności czasu przeżycia od płci, wieku oraz liczby kalorii spożywanych w posiłkach (`meal.cal`). Które zmienne są istotne w tym modelu? Stosując kryterium AIC spróbuj zredukować model.
3. (S) Zbiór danych `ovarian` z pakietu `survival` zawiera informacje na temat czasów przeżycia pacjentów z rakiem jajnika poddanych dwóm terapiom. Zmienna `futime` zawiera informacje o czasach przeżycia, zmienna `fustat` informuje o tym czy czas jest cenzorowany, natomiast zmienna `rx` mówi o rodzaju terapii. Poza tym mamy jeszcze informacje o wieku (`age`) minimalnej obecności choroby po terapii (`resid.ds`) oraz o statusie pacjenta (`ecog.ps`). Przygotuj dane według poniższej formuły.

```
ovarian %>%
  mutate(rx = factor(rx,
                     levels = c('1', '2'),
                     labels = c('A', 'B')),
         resid.ds = factor(resid.ds,
                           levels = c('1', '2'),
                           labels = c('no', 'yes')),
         ecog.ps = factor(ecog.ps,
                          levels = c('1', '2'),
                          labels = c('good', 'bad')),
         age = factor(ifelse(age >=50, 'old', 'young'))),
```

```
futime = Surv(futime, fustat)) -> df
```

- a) Dla tak przygotowanego zbioru danych skonstruuj i narysuj estymator KM ze względu na rodzaj terapii.
- b) Czy krzywe przeżycia różnią się istotnie?
- c) Jakie są średnie czasy przeżycia dla obu terapii?
- d) Skonstruuj pełny model zależności stosując model proporcjonalnych hazardów Coxa.
- e) Czy wszystkie zmienne są istotne w modelu?
- f) Stosując kryterium AIC zreduj model.