

Analiza przeżycia

Laboratorium 7

Alicja Jokiel-Rokita

21 listopada 2020

1 Zadania do sprawozdania 2, część 3

Zadania dotyczą tych samych danych, które opisane były na poprzedniej liście. Poniżej, dla przypomnienia, przedstawiony jest ten opis.

Dane w pliku *CzasDoDializy.csv* dotyczą pacjentów leczonych na niewydolność nerek. Zmienna *Czas* informuje o czasie, w którym zaczęto dializować pacjenta od momentu rozpoznania choroby (w latach), jeżeli zmienna *Obcięcie* wynosi 1, lub czasie obserwacji pacjenta od momentu rozpoznania choroby (w latach), gdy zmienna *Obcięcie* wynosi 0. Pozostałe cztery zmienne dotyczą genotypu pacjenta. Interesuje nas czas od momentu rozpoznania choroby do momentu, gdy dializa jest konieczna do przeprowadzenia. W poniższych zadaniach czas od momentu rozpoznania choroby do dializy pełni rolę “czasu życia”. Oczywiście niektóre dane są cenzurowane i zakładamy, że czas cenzurowania jest niezależny od czasu do wystąpienia zdarzenia. Nie przyjmujemy dodatkowych założeń dotyczących “czasu życia”, tzn. nie przyjmujemy postaci rozkładu tego czasu.

Na podstawie danych *CzasDoDializy.csv* wykonać poniższe zadania.

1. Oszacować punktowo i przedziałowo wartość oczekiwaną i medianę rozkładu czasu od momentu rozpoznania choroby do konieczności przeprowadzenia dializy w całej badanej grupie i w podgrupach ze względu na wartość zmiennej *Arg25Pro*.
2. Uzyskane w punkcie 1. wyniki zamieścić w jednej tabeli.
3. W oparciu o realizację przedziałów ufności, na poziomie ufności 0.95, dla średnich w podgrupach ze względu na wartość zmiennej *Arg25Pro*, sformułować wniosek dotyczący weryfikacji hipotezy o równości średnich w tych podgrupach, na poziomie istotności 0.05.

2 Zadanie dodatkowe

Wyprowadzić wzory na dolną i górną granicę asymptotycznych punktowo przedziałów ufności dla mediany rozkładu czasu życia w oparciu o przekształcenie $g(x) = \log(x)$.