

Wykład "Biostatystyka"
Praca domowa 1
Raport do oddania (mailowo) najpóźniej w dniu 30/03/2020
Prezentacja w dniu 01/04/2020
Punktów do zdobycia: 20

Zbiór „follic_short_cr.csv” zawiera dane dotyczące 517 chorych na chłoniaka leczonych radioterapią i, ewentualnie, chemioterapią. Plik zawiera następujące zmienne:

stnum	- identyfikator chorego
age	- wiek chorego w chwili diagnozy (lata)
clinstg	- stopień zaawansowania choroby (1 - I, 2 - II)
hgb	- poziom hemoglobiny (g/l)
chemo	- wskaźnik leczenia chemioterapią (0 - nie, 1 - tak)
dftime	- czas do wystąpienia nawrotu choroby lub zgonu (lata)
dfstat	- wskaźnik zdarzenia (0 - żył bez nawrotu, 1 - nawrót i/lub zgon)

Interesuje nas, które ze zmiennych charakteryzujących chorego mają wpływ na czas przeżycia bez nawrotu choroby, tzn. czas do wystąpienia nawrotu lub zgonu. W celu udzielenia odpowiedzi na to pytanie, przeprowadź odpowiednią analizę danych wykorzystując metody przedstawione na wykładach 1-2 (szacowanie funkcji przeżycia, testy nieparametryczne). W szczególności:

- Przedstaw wykresy i/lub statystyki opisowe, które pozwolą ocenić wpływ zmiennych na czas przeżycia bez nawrotu choroby.
- Przeprowadź testy, pozwalający na formalną ocenę hipotez dotyczących różnic w funkcjach przeżycia.
- Oceń, czy testy powinny brać pod uwagę wpływ innych czynników niż analizowany. Jeśli tak, przeprowadź odpowiednią analizę.

Analiza może być przeprowadzona przy użyciu SASa lub R. Krótki raport (do czterech stron formatu A4, w tym max. jedna strona na syntaks) może zawierać wydruki z programów, ale MUSI również zawierać opis i interpretację uzyskanych wyników. Raporty oparte tylko o wydruki nie będą brane pod uwagę.

Na podstawie raportu przygotuj krótką prezentację (max. 5 minut) podsumowującą metody analizy i jej wyniki. Możesz zostać poproszony o jej przedstawienie.