

**Wykład "Analiza przeżycia"**  
**Praca domowa 1**  
**Do oddania przed 24/03/2014**  
**Punktów do zdobycia: 20**

**1. (5 pkt)** Dane zamieszczone poniżej przedstawiają czasy przeżycia (obserwacje cenzurowane są podkreślone) dla dwóch grup pacjentów (w sumie 30) z usuniętymi chirurgicznie zmianami czerniaka złośliwego. Pacjenci byli leczeni dwoma rodzajami leczenia immunologicznego:

Szczepionka 1: 4, 5, 8, 10, 17, 17, 17, 19, 24, 34, 34

Szczepionka 2: 7, 8, 8, 8, 8, 8, 8, 11, 11, 12, 12, 15, 16, 18, 21, 23, 24, 25, 27

Odpowiedz na poniższe pytania:

Jaka jest całkowita liczba oczekiwanych zgonów dla każdej z dwóch grup?

Ile wynosi oszacowanie wariancji dla całkowitej liczby zdarzeń w pierwszej z grup?

Ile wynosi wartość statystyki testowej dla testu logrank opartej na oszacowaniu wariancji podanym powyżej?

Czy wartość statystyki jest statystycznie istotna na poziomie istotności 0.05?

Ile wynosi wartość statystyki testowej dla testu logrank wyznaczona przy pomocy „prostszych obliczeń”?

**2. (5 pkt)** Dane z punktu 1 powyżej zostały podzielone na warstwy odpowiadające wiekowi pacjentów (obserwacje cenzurowane są podkreślone):

Wiek 21-40 lat:

Szczepionka 1: 8, 17, 17, 19, 24, 34

Szczepionka 2: 7, 8, 8, 12, 16, 18, 21, 24, 27

Wiek 41-60 lat:

Szczepionka 1: 4, 17, 34

Szczepionka 2: 8, 8, 8, 11, 12, 15, 23

Wiek 61 lat i więcej:

Szczepionka 1: 5, 10

Szczepionka 2: 8, 11, 25

Ile wynosi oczekiwana liczba zgonów dla każdej z grup w każdej z warstw?

Ile wynoszą sumaryczne (sumowane po warstwach) oczekiwane liczby zgonów dla każdej z grup?

Ile wynosi wartość statystyki testowej dla warstwowego (ze względu na wiek) testu logrank wyznaczona przy pomocy „prostszych obliczeń”?

Czy wartość statystyki jest statystycznie istotna na poziomie istotności 0.05?

**3. (10 pkt)** Plik "cyfra\_short.csv" zawiera obserwacje czasu przeżycia (*survtime*, w miesiącach) i wskaźnik zdarzeń (*survind*, 1 dla zgonu, 0 dla obserwacji cenzurowanej) dla 94 pacjentów z operowalnym nie-drobnokomórkowym rakiem płuca (non-small cell lung cancer, NSCLC). Głównym obiektem zainteresowania jest zmienna *level*, która wskazuje pacjentów z niskimi (0) i wysokimi (0) poziomami markera CYFRA-21. Uważa się, że marker ten różnicuje prognozę pacjentów z NSCLC. Dodatkowo plik zawiera zmienne *stage* (stopień zaawansowania nowotworu TNM; 1 dla TNM I, 2 dla TNM II, i 3 dla TNM IIIA) oraz *histpat* (typ histologiczny nowotworu: *histpat*=0 dla gruczolakoraka, 1 dla raka wielkokomórkowego, 2 dla raka płaskonabłonkowego).

Przeprowadź analizę danych. W szczególności:

- Przedstaw wykresy i/lub statystyki opisowe, które pozwolą ocenić wpływ zmiennych na czas przeżycia. Czy czas przeżycia chorych z wyższymi poziomami markera był średnio dłuższy czy krótszy od czasu przeżycia chorych z niższymi poziomami markera?

- Przeprowadź test, pozwalający na formalną ocenę hipotez dotyczących różnic w funkcjach przeżycia dla grup chorych różniących się poziomem markera.
- Oceń, czy test powinien brać pod uwagę wpływ stopnia zaawansowania nowotworu i/lub typu histologicznego. Jeśli tak, przeprowadź odpowiednią analizę.

Plik można wczytać do R przy użyciu polecenia (ze zmodyfikowaną ścieżką dostępu)

```
> cyfra <- read.csv("c:/cyfra_short.csv")
```

W SASie można posłużyć się poleceniem **Import data** z menu **File**. Należy wybrać format .csv jako „*data source*”, wskazać lokalizację pliku, library i nazwę pod jaką dane mają być zachowane (standardowo SAS proponuje tymczasową library WORK; ale można najpierw zdefiniować libname dla jakiegoś katalogu i wybrać to libname; wówczas plik danych w formacie SASa zostanie zapisany na stałe w tym katalogu) i polecić wykonanie importu.

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